Plan du cours

1. Introduction
2. Statistique descriptive – séries univariées
3. Calcul des probabilités
4. Arbres de décision
5. Variables aléatoires et lois de probabilité
6. Statistique descriptive – séries bivariées
7. Méthodes de prévision

2. Stat. Descriptive - 1 dim

- **Objectif** : Résumer les caractéristiques d'un (grand) ensemble de données. Mettre en évidence les points importants.
- **Tableaux** : Première étape. Tri et regroupement des observations. Distribution de fréquences.
- **Graphiques** : Visualisation des données. Lignes, barres, histogrammes, diagrammes en secteurs, …
- **Mesures** : Synthèse des données en quelques grandeurs représentatives.
  - Quel est l'ordre de grandeur des valeurs observées ?
    Paramètres de position (moyenne, médiane, …)
  - Y a-t-il de grands écarts entre les valeurs observées ?
    Paramètres de dispersion (variance, écart-type, …)

Données brutes

$$\{x_1, x_2, \cdots, x_n\} = \{x_i; i = 1, \cdots, n\}$$

- Données telles que recueillies
  - n observations d'une variable x : série statistique univariée
  - Exemple 2 : Ages de 100 employés d'une entreprise (échantillon...)

60	39	23	30	29	26	29	41	40	32
63	22	32	52	46	35	25	28	33	33
20	25	42	34	29	43	41	31	30	36
58	21	24	55	51	28	18	40	44	38
32	21	30	31	25	49	31	26	33	36
43	34	35	22	33	38	34	34	33	34
23	26	57	23	26	36	39	31	35	34
34	51	40	50	35	45	28	36	32	39
26	48	17	45	45	25	25	30	36	30
43	25	27	21	53	25	38	33	37	33

Série ordonnée

- Si x est ordinale ou quantitative

$$\left\{x_{(1)}, x_{(2)}, \ldots, x_{(n)}\right\} \text{ avec } x_{(i)} \leq x_{(j)} \text{ si } i \leq j$$

17	18	20	21	21	21	22	22	23	23
23	24	25	25	25	25	25	25	25	26
26	26	26	26	27	28	28	28	29	29
29	30	30	30	30	30	31	31	31	31
32	32	32	32	33	33	33	33	33	33
33	34	34	34	34	34	34	34	35	35
35	35	36	36	36	36	36	37	38	38
38	39	39	39	40	40	40	41	41	42
43	43	43	44	45	45	45	46	48	49
50	51	51	52	53	55	57	58	60	63

Distribution observée

Age	Effectif
15-19	2
20-24	10
25-29	19
30-34	27
35-39	16
40-44	10
45-49	6
50-54	5
55-59	3
60-64	2
Total	100

- Données regroupées en classes (intervalles).
- Effectif d'une classe = nombre d'observations dans cette classe.

Exemple 3

- Nb de voitures dans 10 familles de même taille.

i	1	2	3	4	5	6	7	8	9	10
x_i	2	0	1	1	2	3	1	1	1	0

$$\boxed{n=10}$$

- Série observée :

$$\{x_i\}=\{2,0,1,1,2,3,1,1,1,0\}$$

- Série ordonnée :

$$\{x_{(i)}\}=\{0,0,1,1,1,1,1,2,2,3\}$$

Exemple 3

- Distribution observée :

Valeurs observées → x_j

Effectifs associés → n_j

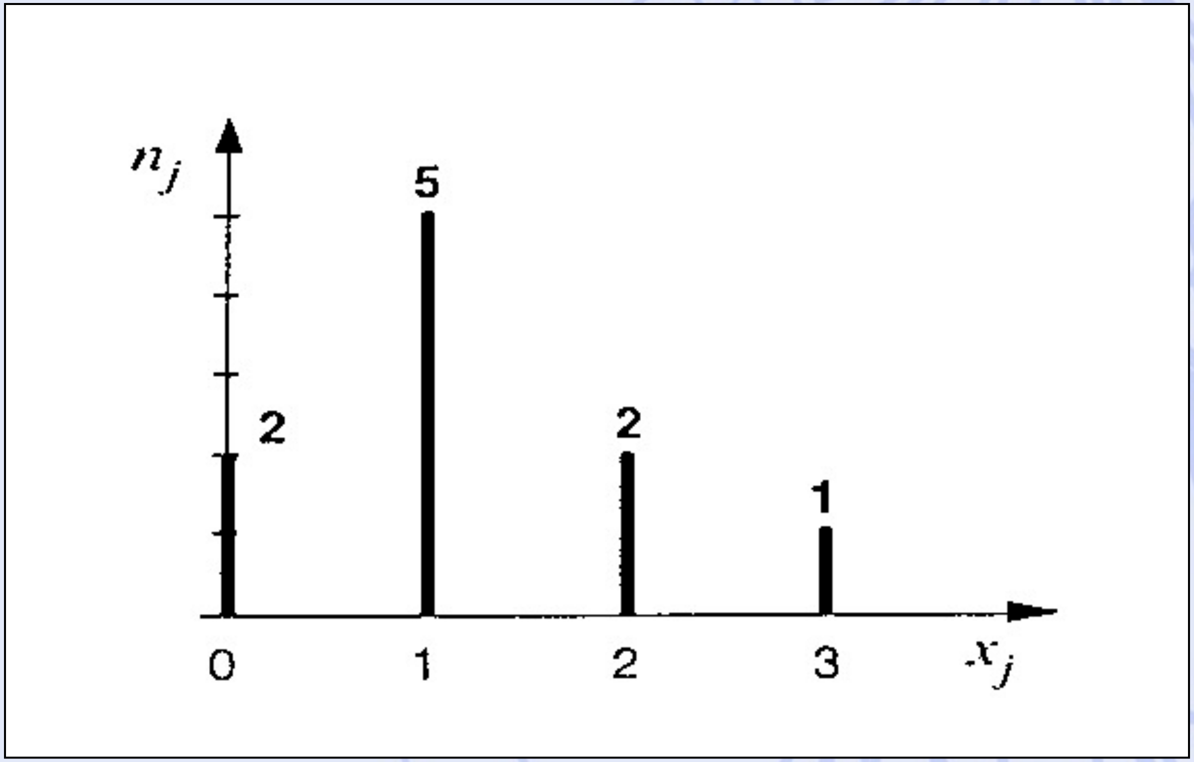
x_j	0	1	2	3
n_j	2	5	2	1

$$j = 1, 2, 3, 4 \quad J = 4 = \text{ nb de valeurs distinctes}$$

$$n_1 + n_2 + \ldots + n_J = \sum_{j=1}^{J} n_j = n$$

Représentations graphiques

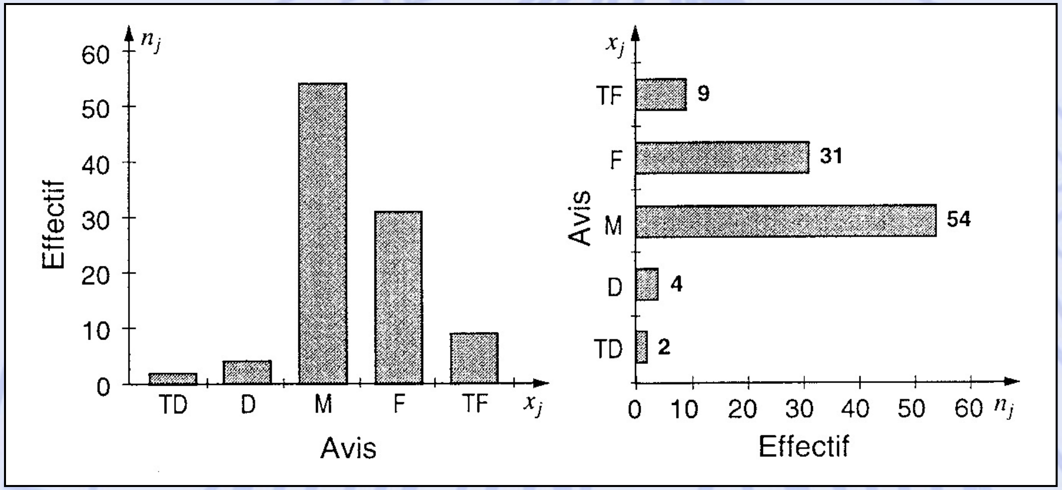
- Diagramme en bâtons : exemple 3

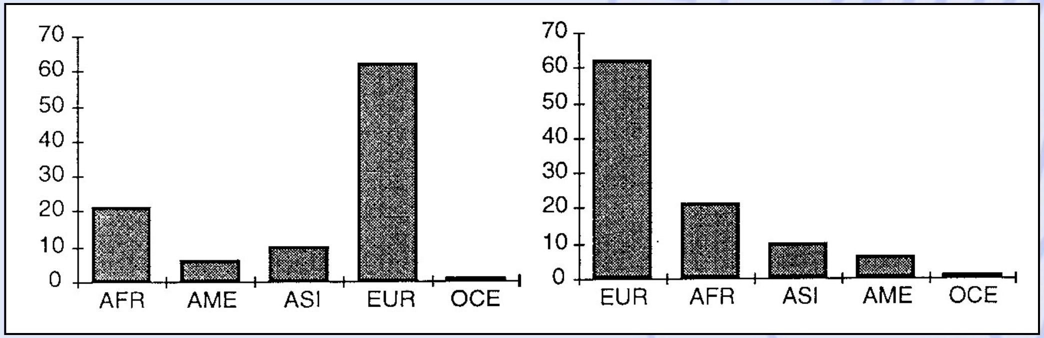


Représentations graphiques

- Diagramme en barres : pour variables qualitatives

Cas ordinal
(avis pédagogiques)

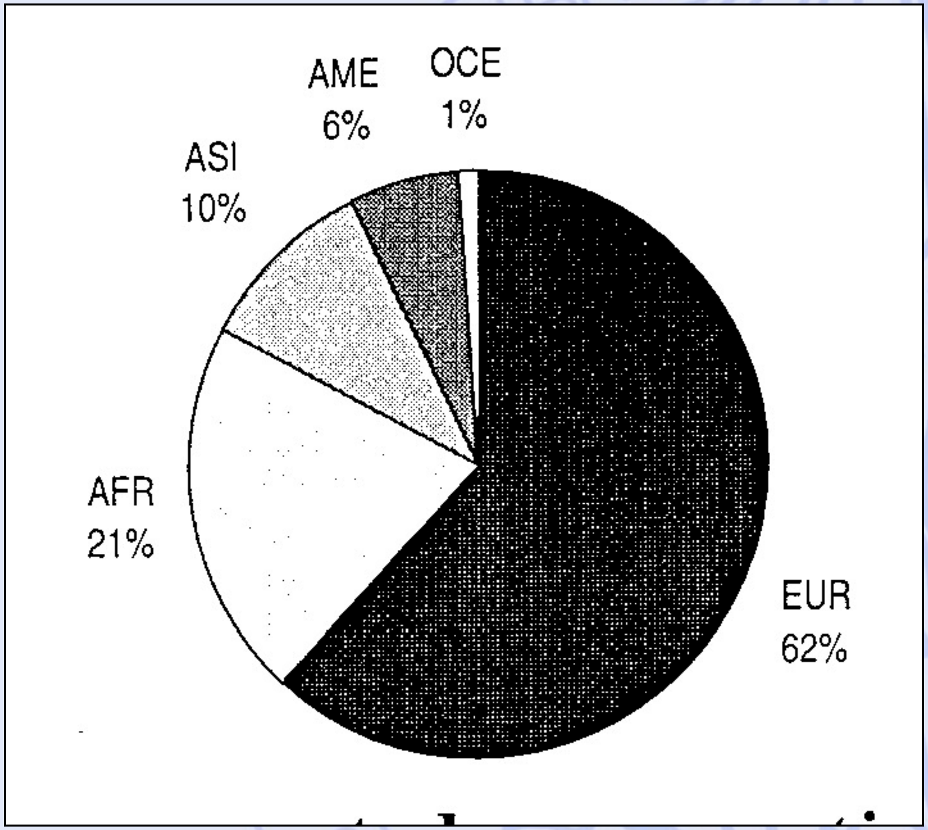




Cas nominal
(continents)

Représentations graphiques

- Diagramme en secteurs (tarte, camembert):



Fréquences

- Définition : $f_j = \dfrac{n_j}{n} \quad j = 1, 2, \ldots, J$
- Exemple 3 : distribution de fréquences

x_j	n_j	f_j
0	2	0.2
1	5	0.5
2	2	0.2
3	1	0.1

$$n = 10$$

$$\sum_{j=1}^{J} f_j = 1$$

Effectifs et fréquences cumulés

- Effectifs cumulés :

$$\text{nombre d'observations} \leq x_j$$

$$N_j = n_1 + n_2 + \ldots + n_j$$

- Fréquences cumulées :

$$\text{pourcentage d'observations} \leq x_j$$

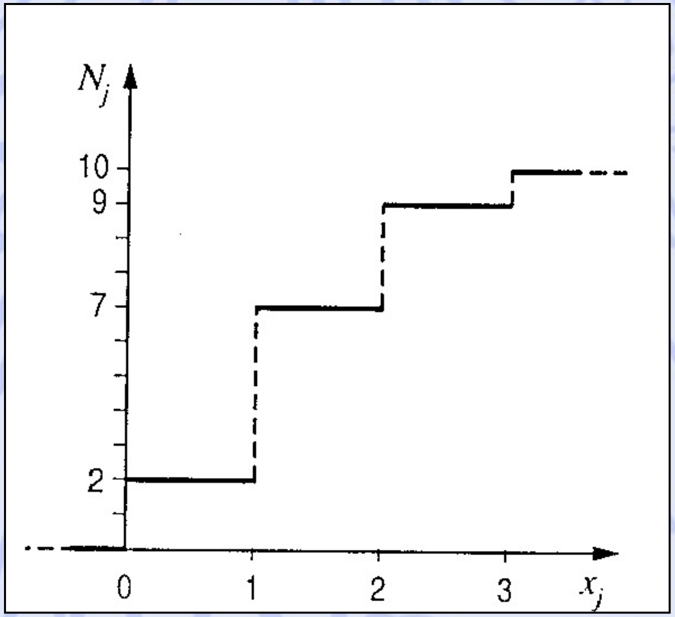
$$F_j = \frac{N_j}{n}$$

Effectifs et fréquences cumulés

- Exemple 3 :

x_j	n_j	f_j	N_j	F_j
0	2	0.2	2	0.2
1	5	0.5	7	0.7
2	2	0.2	9	0.9
3	1	0.1	10	1.0

Courbe cumulative



Effectifs et fréquences cumulés à droite

- Effectifs cumulés à droite :

$$\text{nombre d'observations } \geq x_j$$

$$N_j^* = n_j + n_{j+1} + \ldots + n_J$$

- Fréquences cumulées à droite :

$$\text{pourcentage d'observations } \geq x_j$$

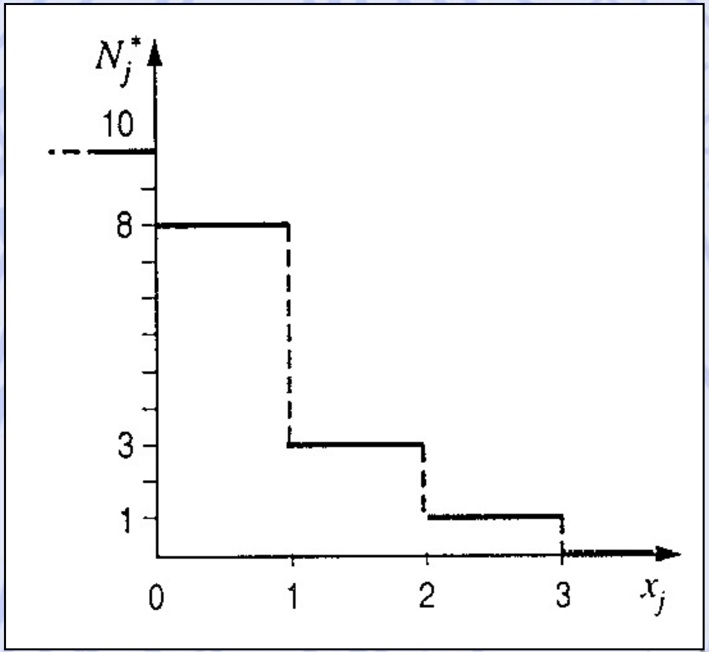
$$F_j^* = \frac{N_j^*}{n}$$

Effectifs et fréquences cumulés à droite

- Exemple 3 :

x_j	n_j	f_j	N_j^*	F_j^*
0	2	0.2	10	1.0
1	5	0.5	8	0.8
2	2	0.2	3	0.3
3	1	0.1	1	0.1

Courbe cumulative à droite



Distribution groupée

- Cf. exemple 2 :
  - Grand nombre de valeurs distinctes,
  - Petits effectifs.
- Regroupements des valeurs en classes

Exemple 4

- Série de 175 tailles

176	175	185	176	190	163	185
166	152	178	172	178	190	176
177	160	183	186	182	154	172
150	168	174	180	173	174	157
181	171	181	192	188	171	176
169	184	168	179	185	189	165
175	170	176	184	189	176	171
173	189	177	170	165	182	174
184	178	161	150	175	171	194
162	187	166	170	176	183	175
169	198	185	171	184	177	166
171	180	158	177	155	184	195
193	176	192	169	181	178	180
154	155	196	175	175	190	191
170	165	151	191	177	153	180
155	177	187	180	169	198	156
186	164	178	187	183	175	183
163	182	177	183	170	169	182
184	170	190	164	186	164	160
176	175	175	185	169	170	174
180	173	162	161	196	181	181
188	182	181	177	179	177	179
178	172	163	165	198	167	180
169	184	172	167	151	183	161
177	175	168	174	197	179	179

150	160	165	169	172	175	177	178	181	184	187	192
150	161	165	170	172	175	177	179	181	184	187	193
151	161	166	170	172	175	177	179	181	184	187	194
151	161	166	170	173	175	177	179	181	184	188	195
152	162	166	170	173	175	177	179	182	184	188	196
153	162	167	170	173	175	177	179	182	184	189	196
154	163	167	170	174	176	177	180	182	184	189	197
154	163	168	170	174	176	177	180	182	185	189	198
155	163	168	171	174	176	177	180	182	185	190	198
155	164	168	171	174	176	177	180	183	185	190	198
155	164	169	171	174	176	178	180	183	185	190	
156	164	169	171	175	176	178	180	183	185	190	
157	165	169	171	175	176	178	180	183	186	191	
158	165	169	171	175	176	178	181	183	186	191	
160	165	169	172	175	176	178	181	183	186	192	

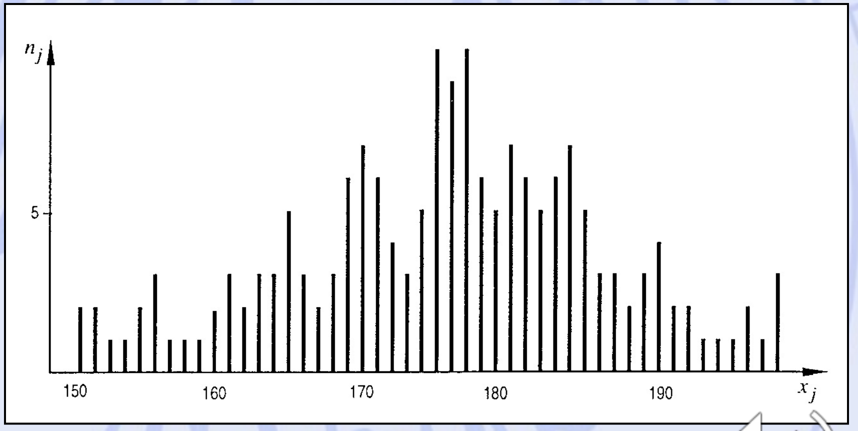


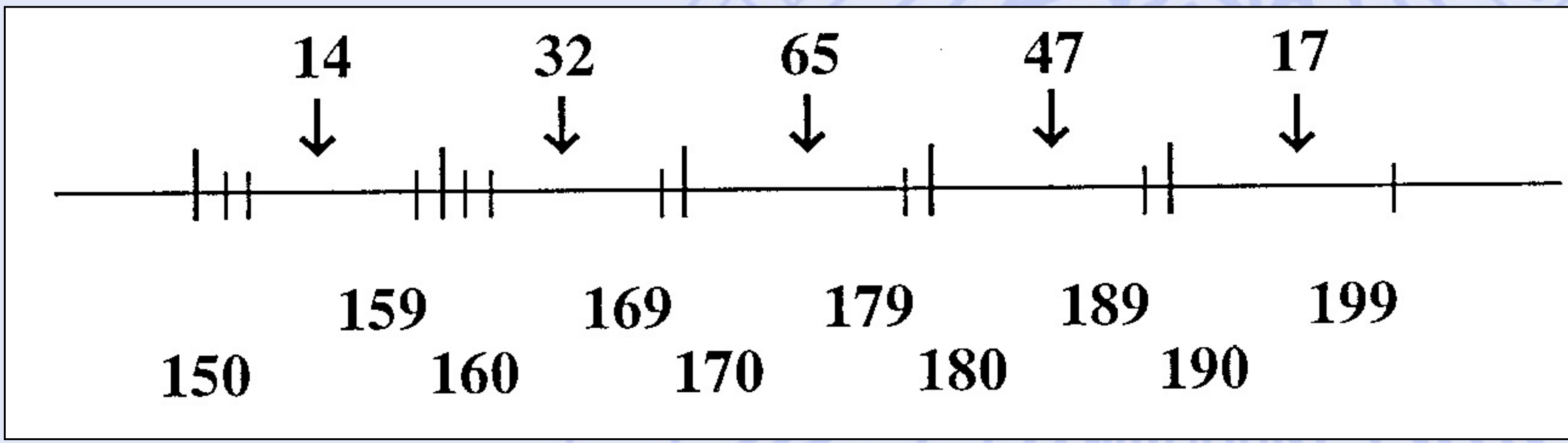
Diagramme en tiges et feuilles

- « Stem and leaf display »

15	04525810514376
16	69253908548162389415759939475601
17	6753106875108670532584678752629001757483565709416178507962614 5499
18	14640849702245315716040735285941369234135003210
19	38260210687008451

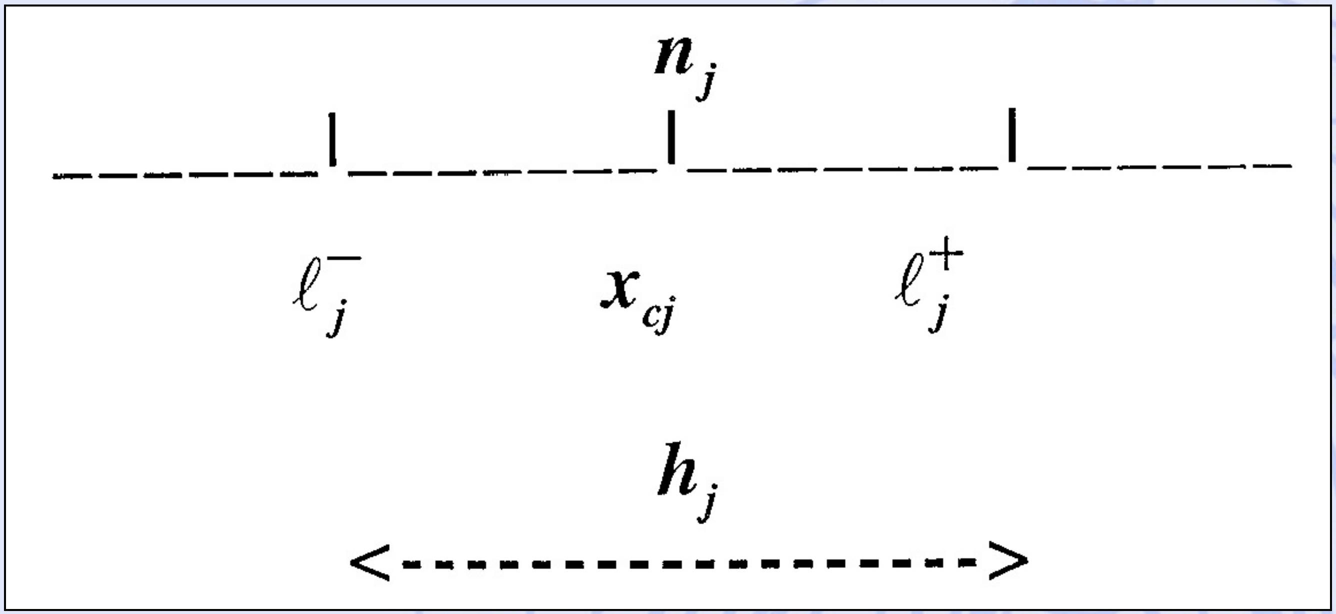
Groupement en classes

- Intervalles disjoints
- Recouvrir toutes les valeurs observées



Groupement en classes

- J classes : $$j = 1, 2, \ldots, J$$



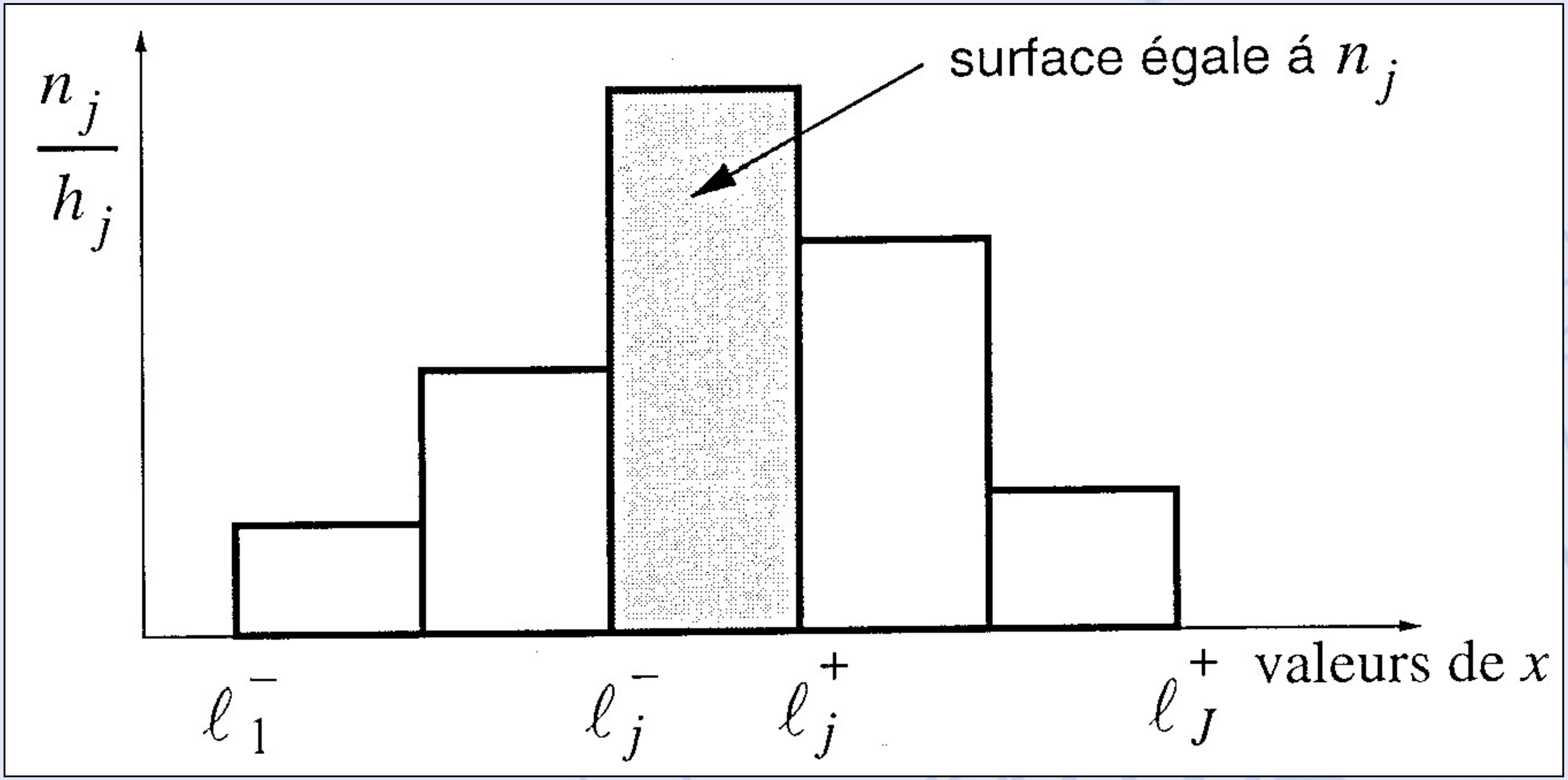
- J = ? Règle de Sturges :

$$J \approx 1 + \frac{10}{3} \log_{10} n$$

Distribution groupée

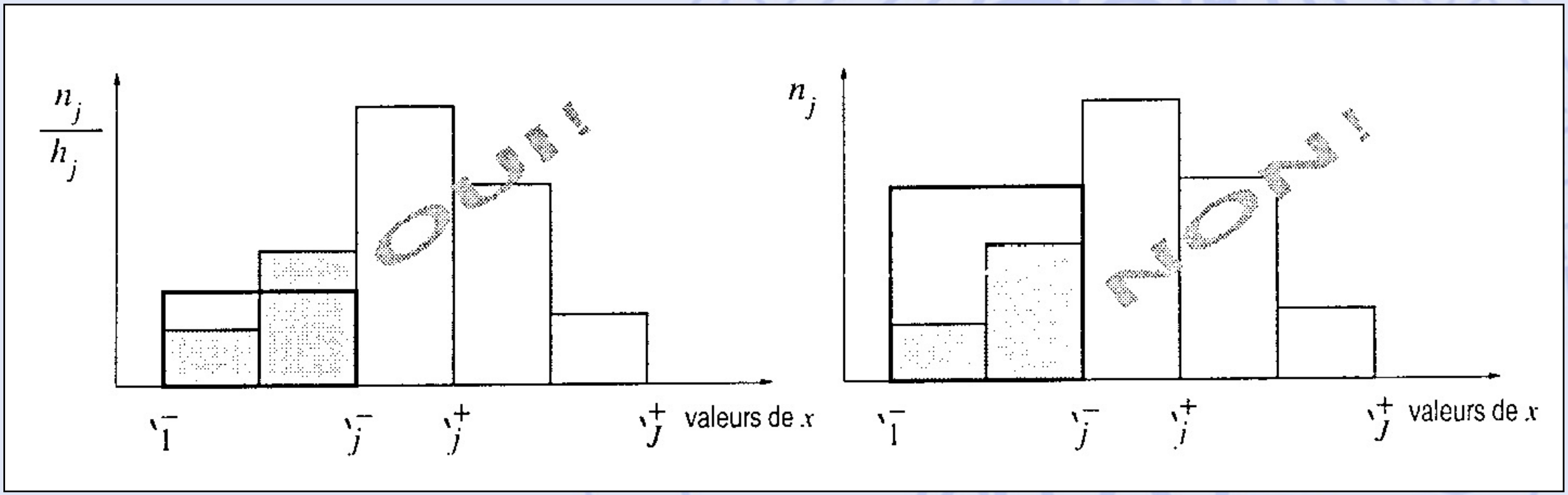
$(\ell_j^- ; \ell_j^+)$	x_{cj}	n_j	f_j	N_j	F_j	N_j^*	F_j^*
149.5-159.5	154.5	14	0.080	14	0.080	175	1
159.5-169.5	164.5	32	0.183	46	0.263	161	0.920
169.5-179.5	174.5	65	0.371	111	0.634	129	0.737
179.5-189.5	184.5	47	0.269	158	0.903	64	0.366
189.5-199.5	194.5	17	0.097	175	1	17	0.097
		n=175	1				

Histogramme des effectifs

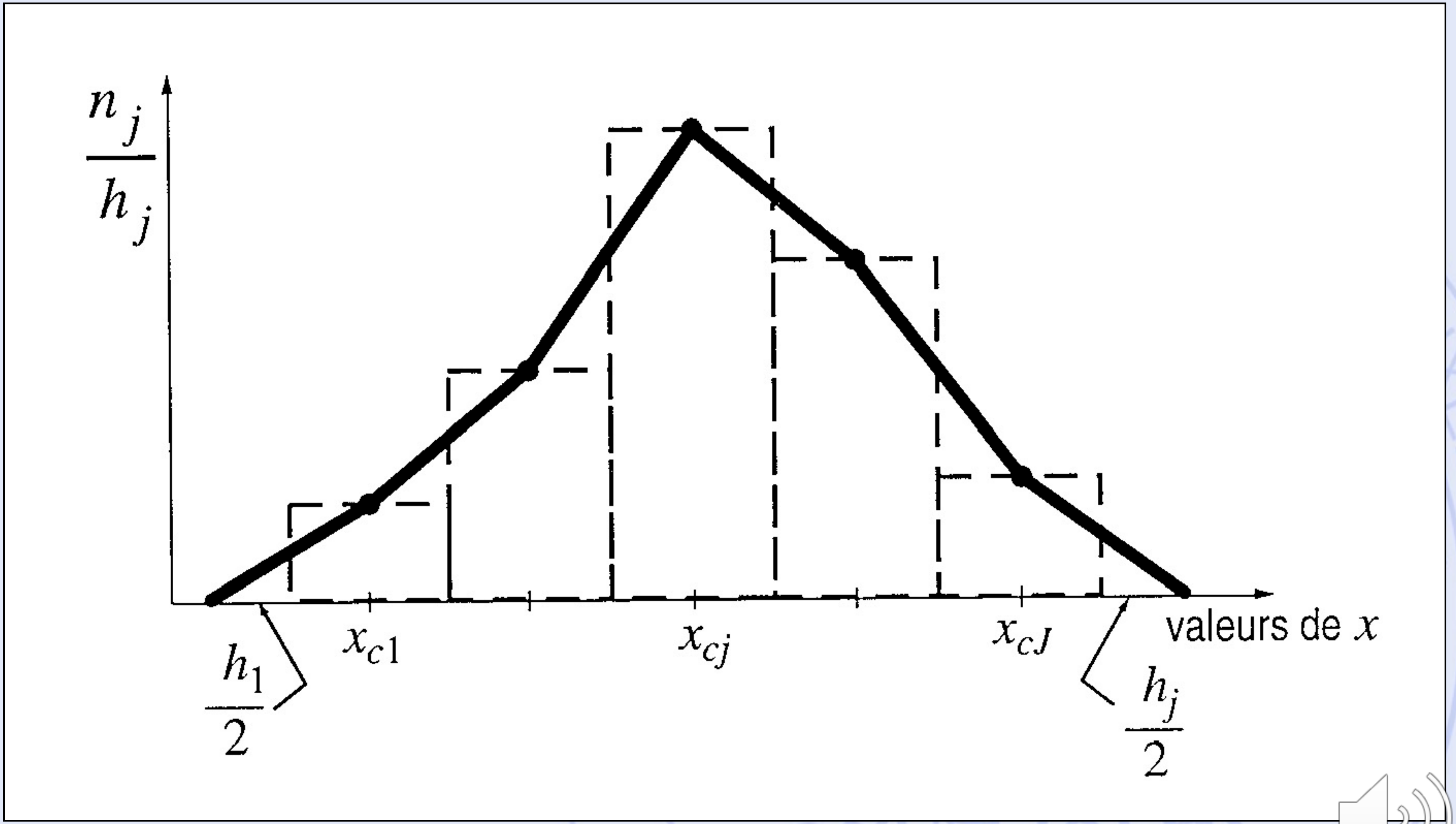


Histogramme des effectifs

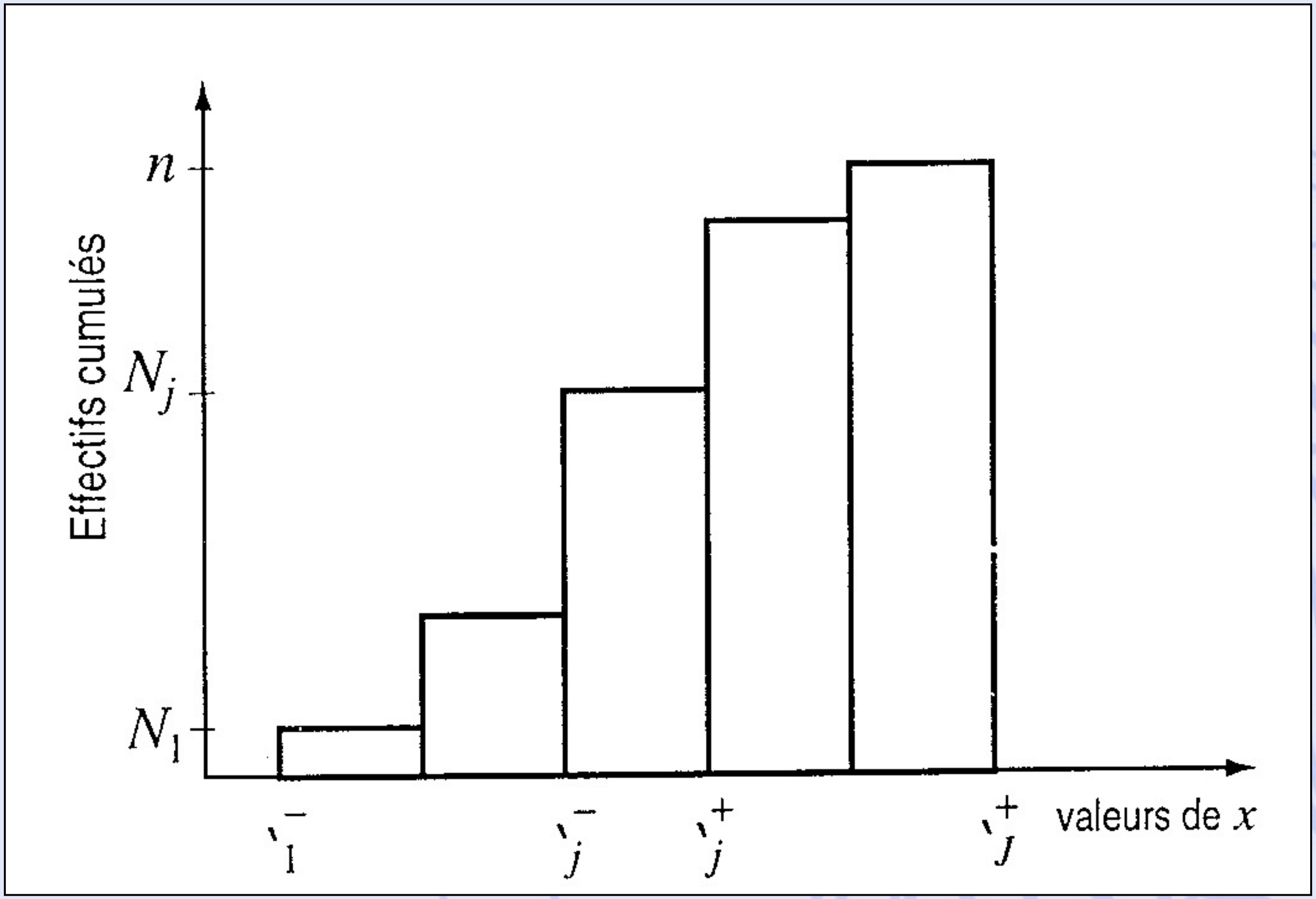
- Attention à l'unité sur l'axe vertical !
- Cas de classes de longueurs différentes.



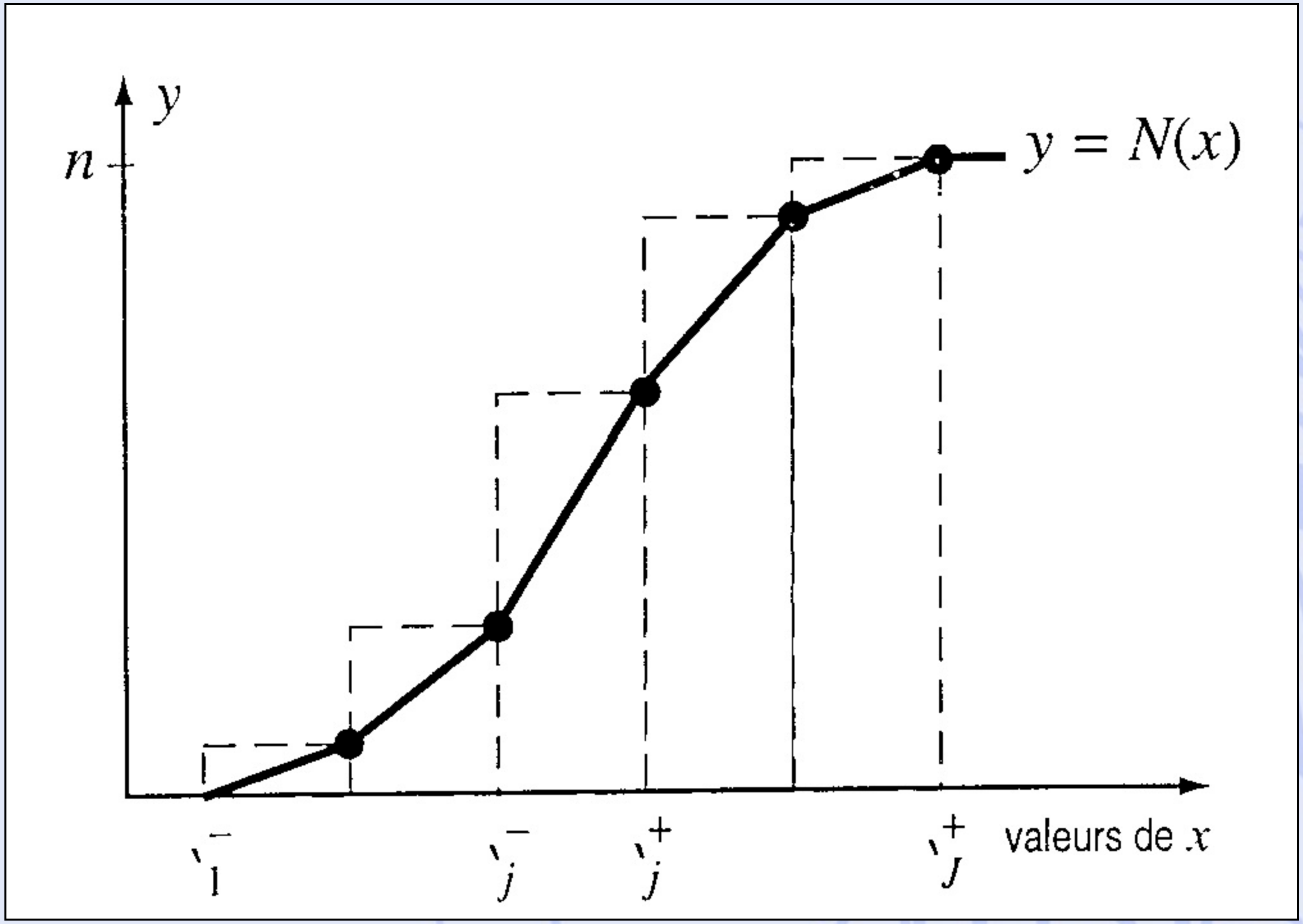
Polygone des effectifs



Histogramme des effectifs cumulés



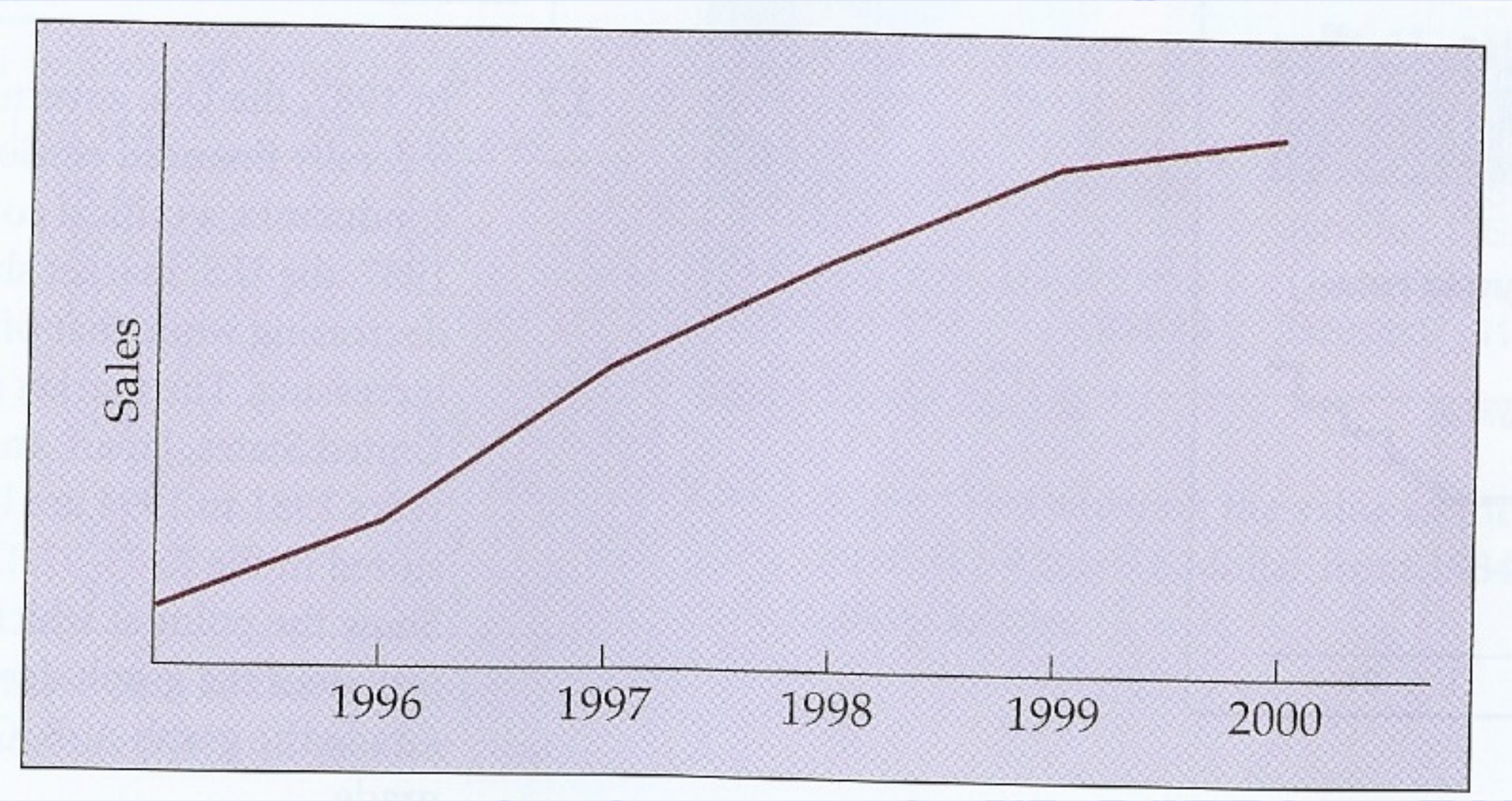
Courbe cumulative



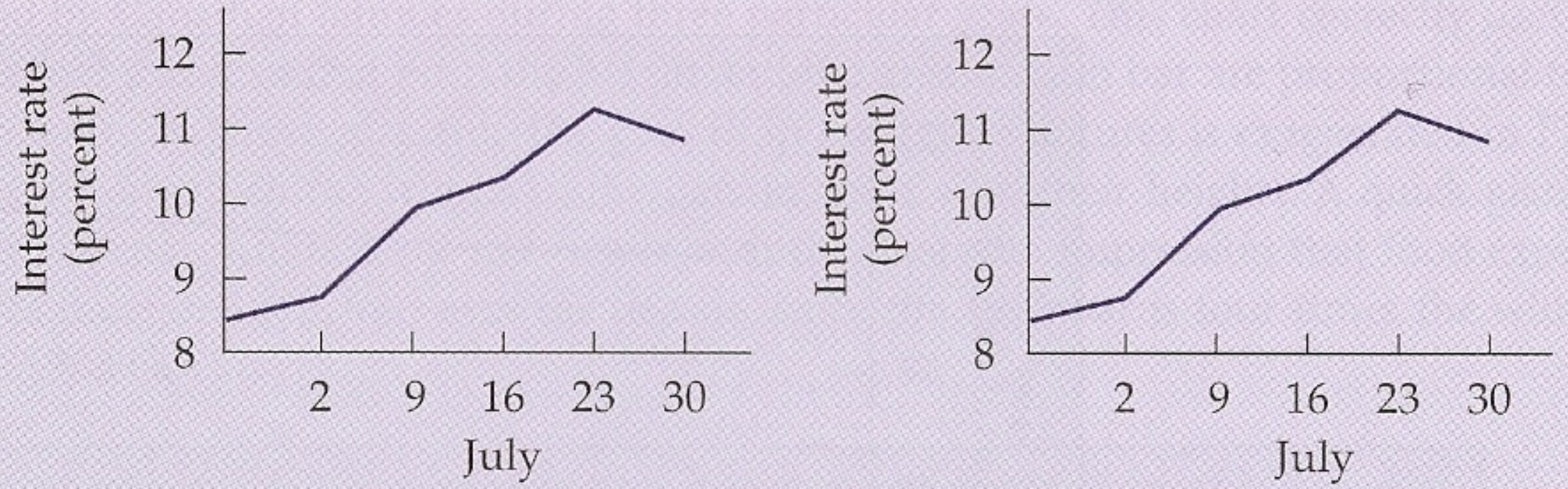
Utilisation de graphiques

- Inclure informations utiles.
- Eviter éléments inutiles !
- Préférer la simplicité à la sophistication.
- Choix des unités et des axes.
- …

Graphiques !



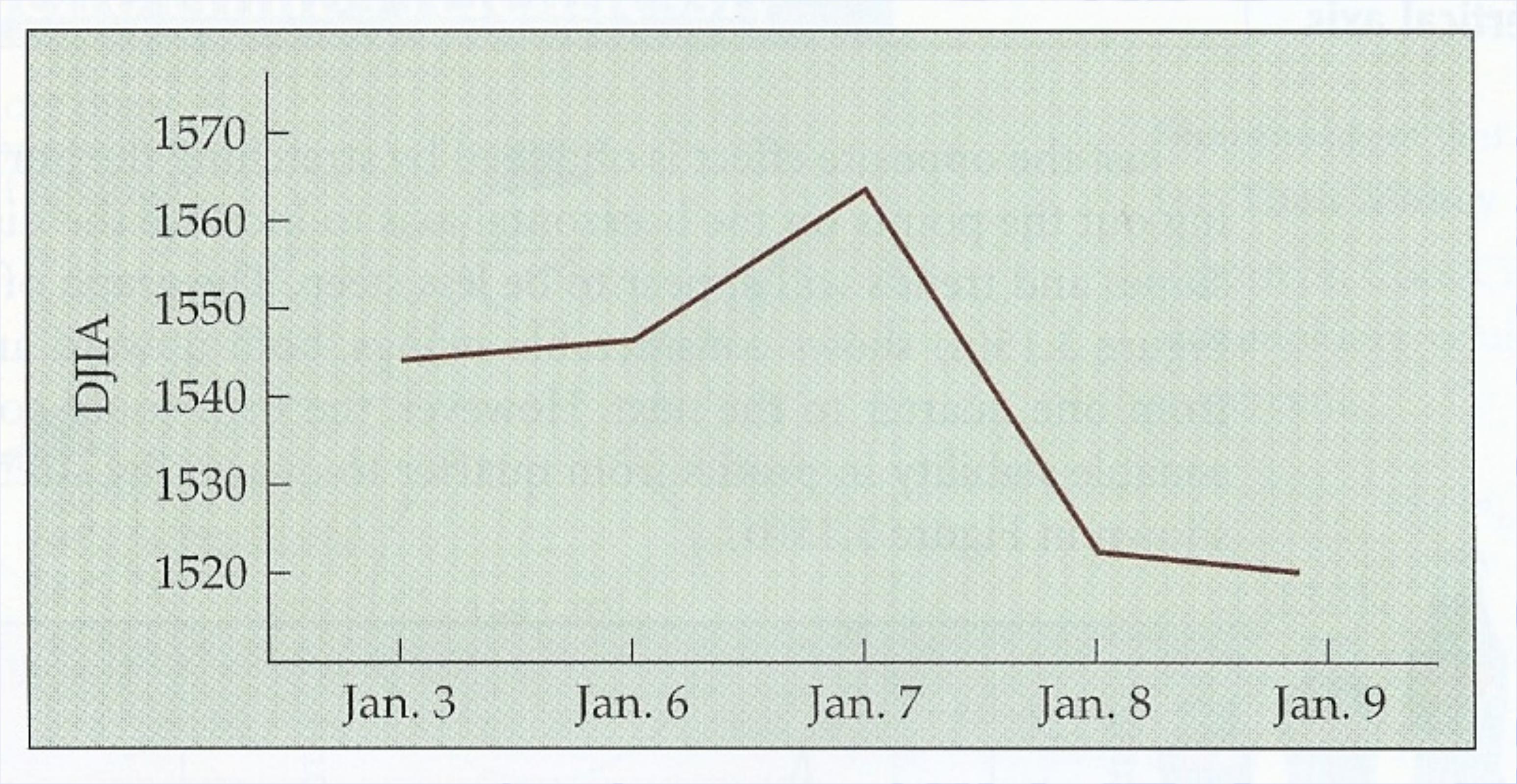
Graphiques !



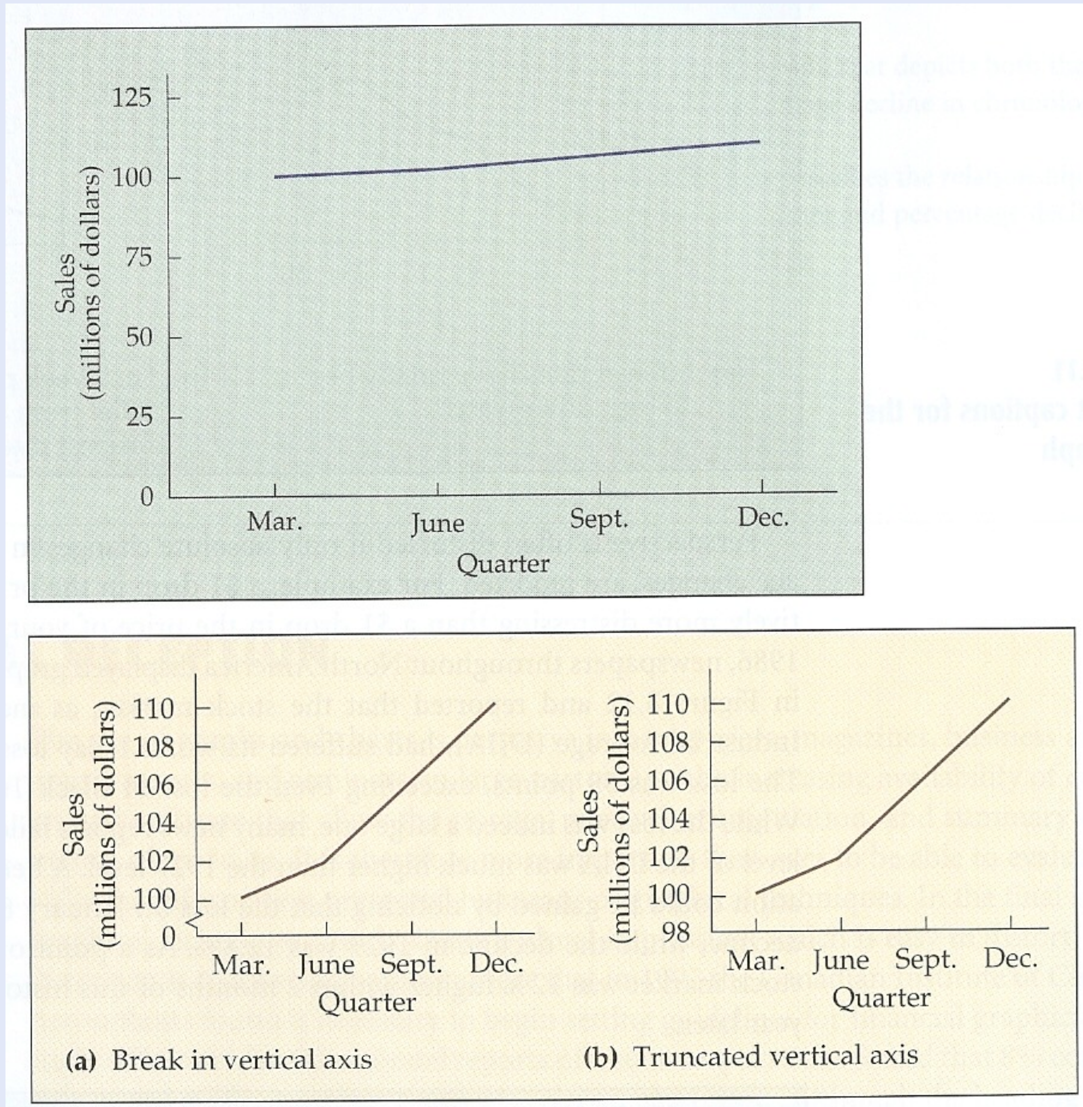
(a) Interest rates have finally begun to turn downward.

(b) Last week provided temporary relief from the upward trend in interest rates.

Graphiques !



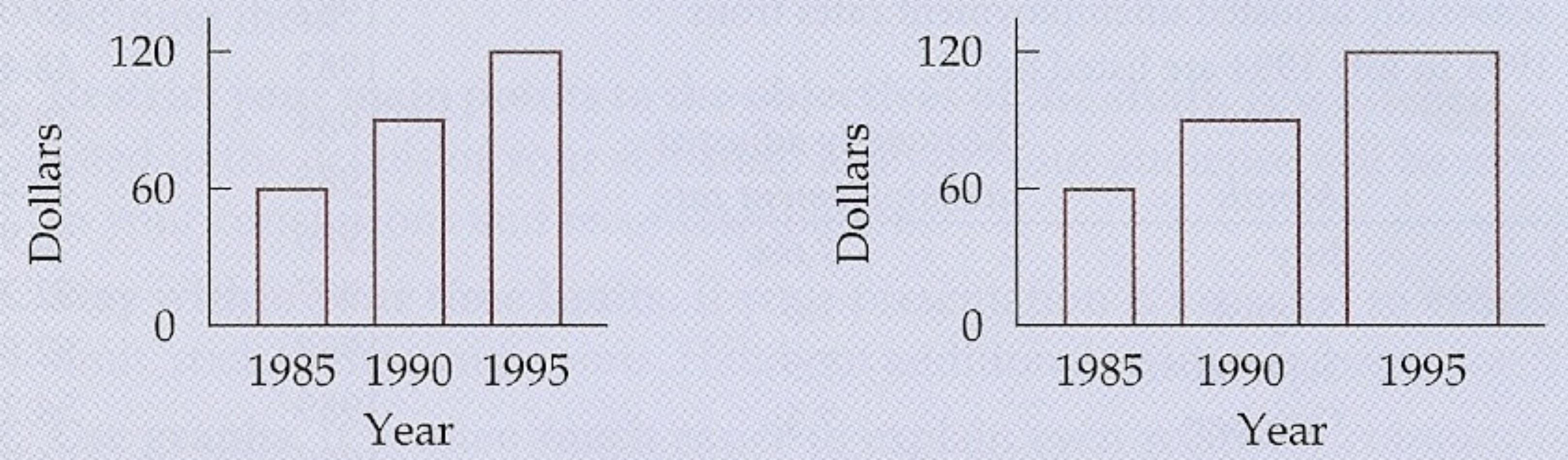
Graphiques !



(a) Break in vertical axis

(b) Truncated vertical axis

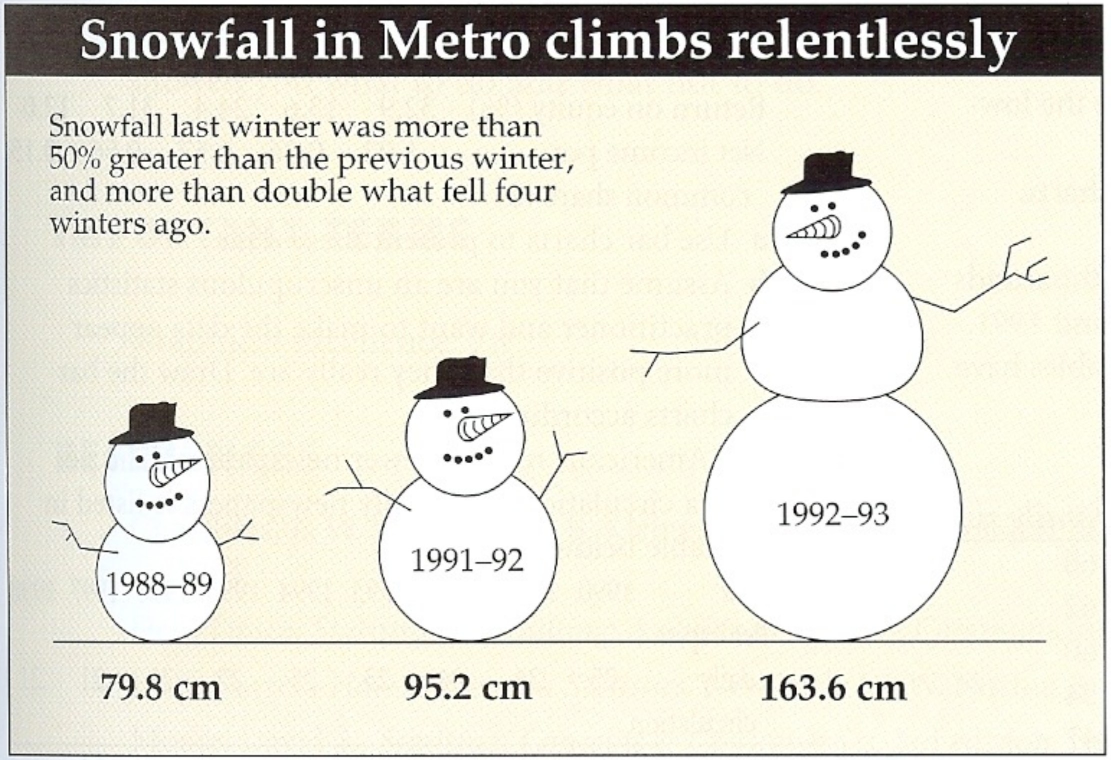
Graphiques !

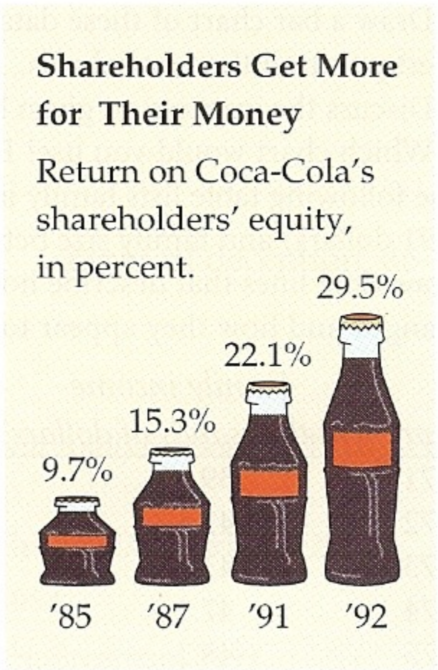


(a) Correct bar chart

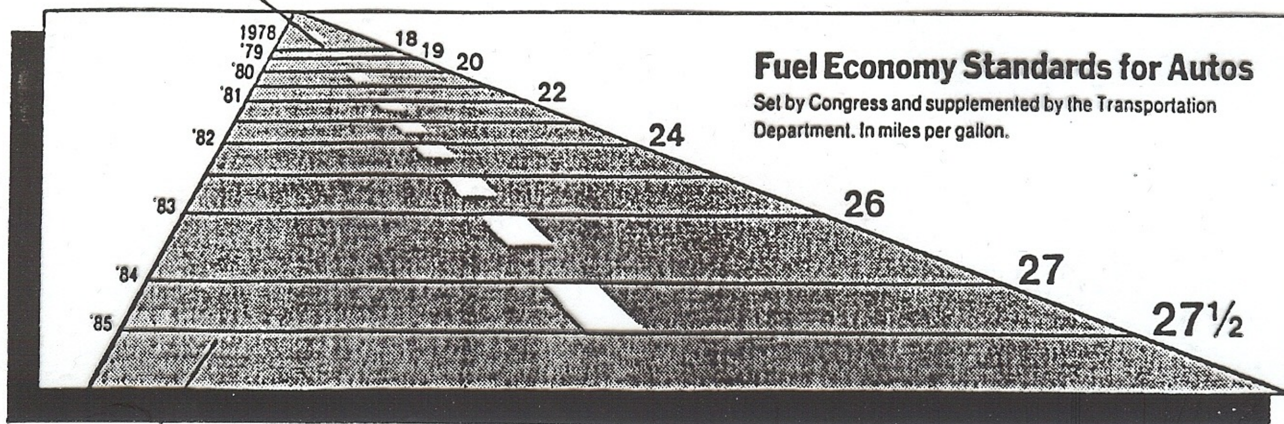
(b) Increasing bar widths to create distortion

Graphiques !


Snowfall in Metro climbs relentlessly
Snowfall last winter was more than 50% greater than the previous winter, and more than double what fell four winters ago.
1988–89
1991–92
1992–93
79.8 cm
95.2 cm
163.6 cm

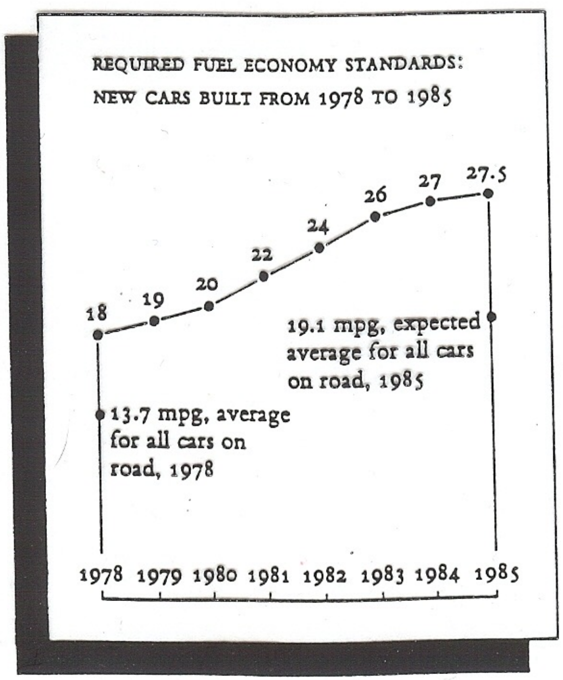

Shareholders Get More for Their Money
Return on Coca-Cola's shareholders' equity, in percent.
9.7%
15.3%
22.1%
29.5%
'85
'87
'91
'92

This line, representing 18 miles per gallon in 1978, is 0.6 inches long.



This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

New York Times, August 9, 1978, p. D-2.



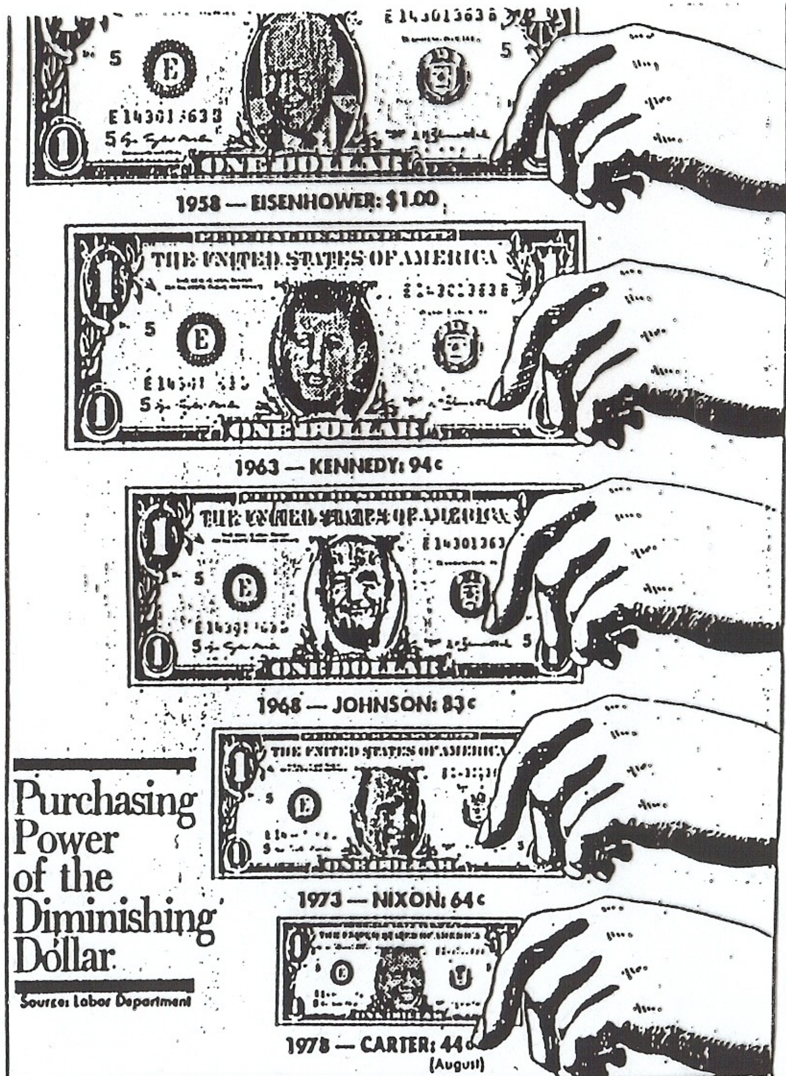


Figure 9. An example of how to goose up the effect by squaring the eyeball (© 1978, The Washington Post).

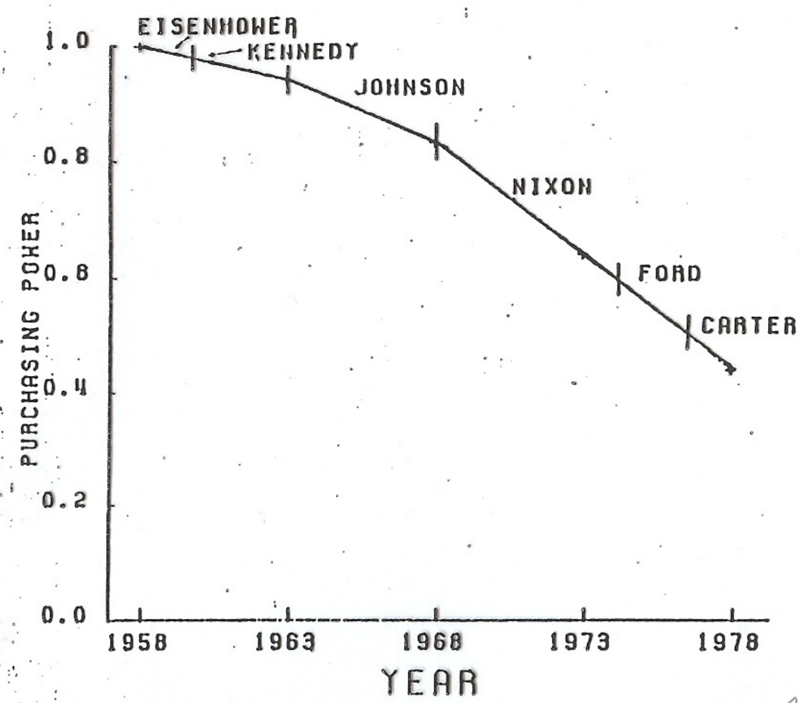


Figure 10. The data in Figure 9 as an unadorned line chart (fr... Wainer, 1980).

THE NEW YORK TIMES, SUNDAY, AUGUST 2, 1981

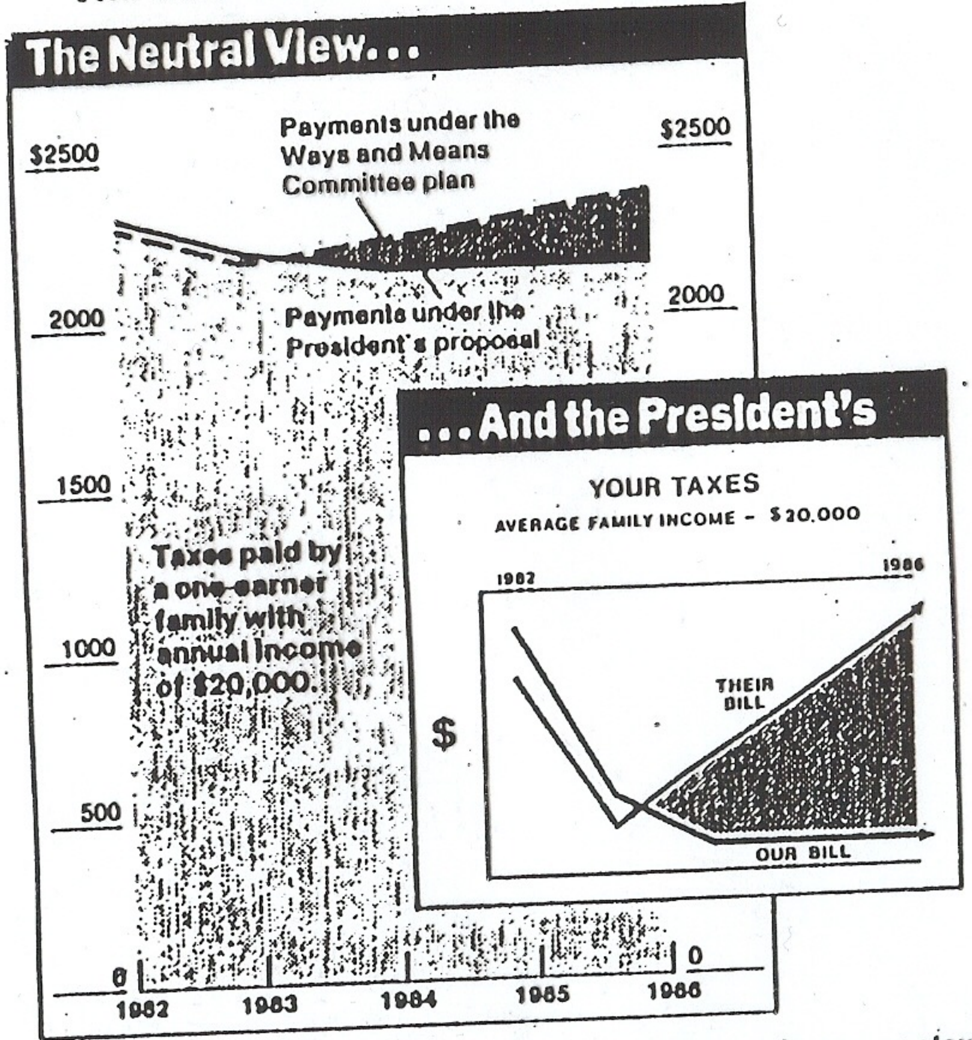


Figure 11. The White House showing neither scale nor context (© 1981, The New York Times, reprinted with permission).

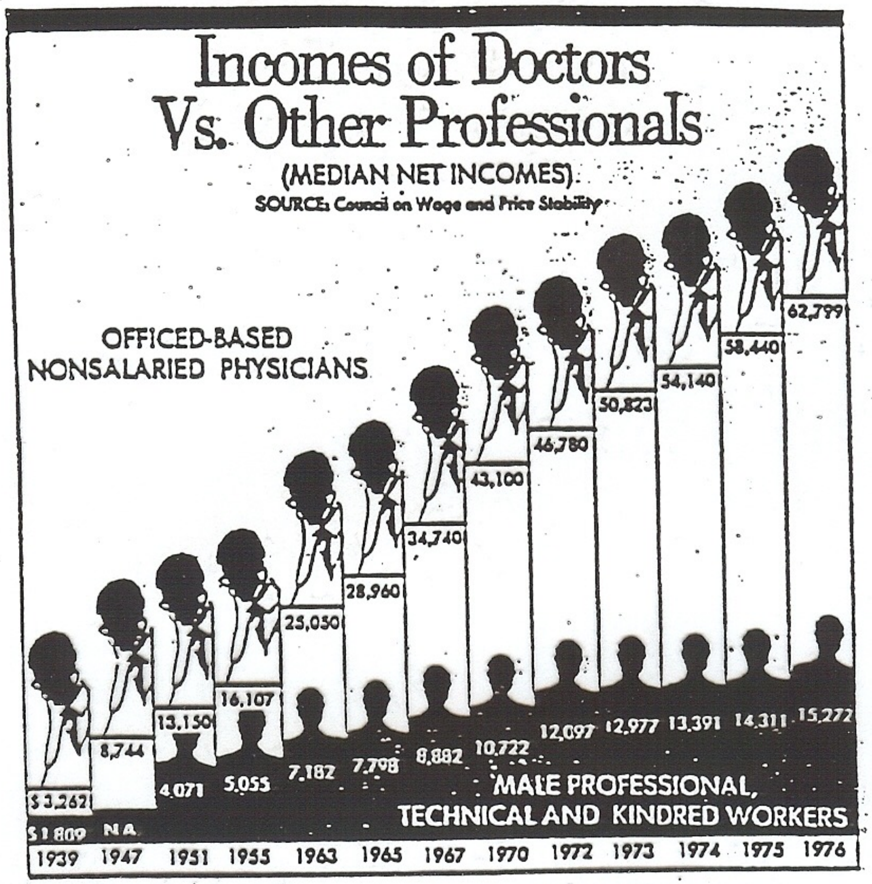


Figure 13. Changing scale in mid-axis to make exponential growth linear (© The Washington Post).

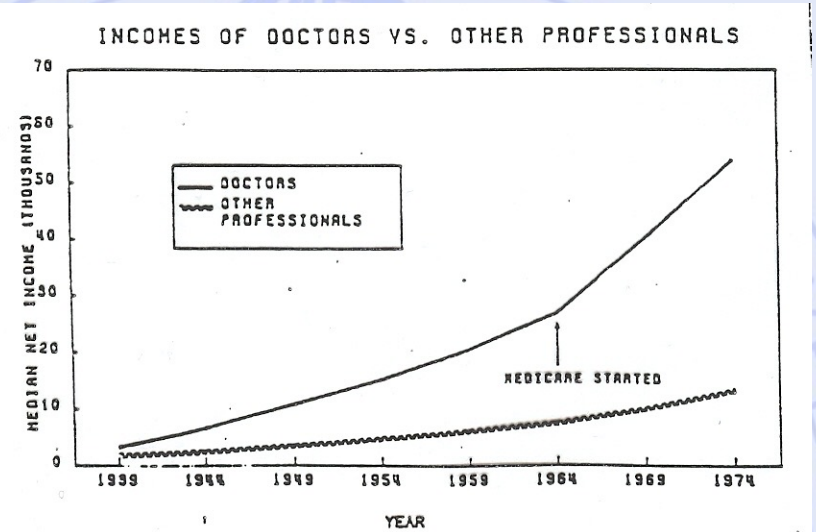


Figure 14. Data from Figure 13 redone with linear scale (from Wainer 1980).

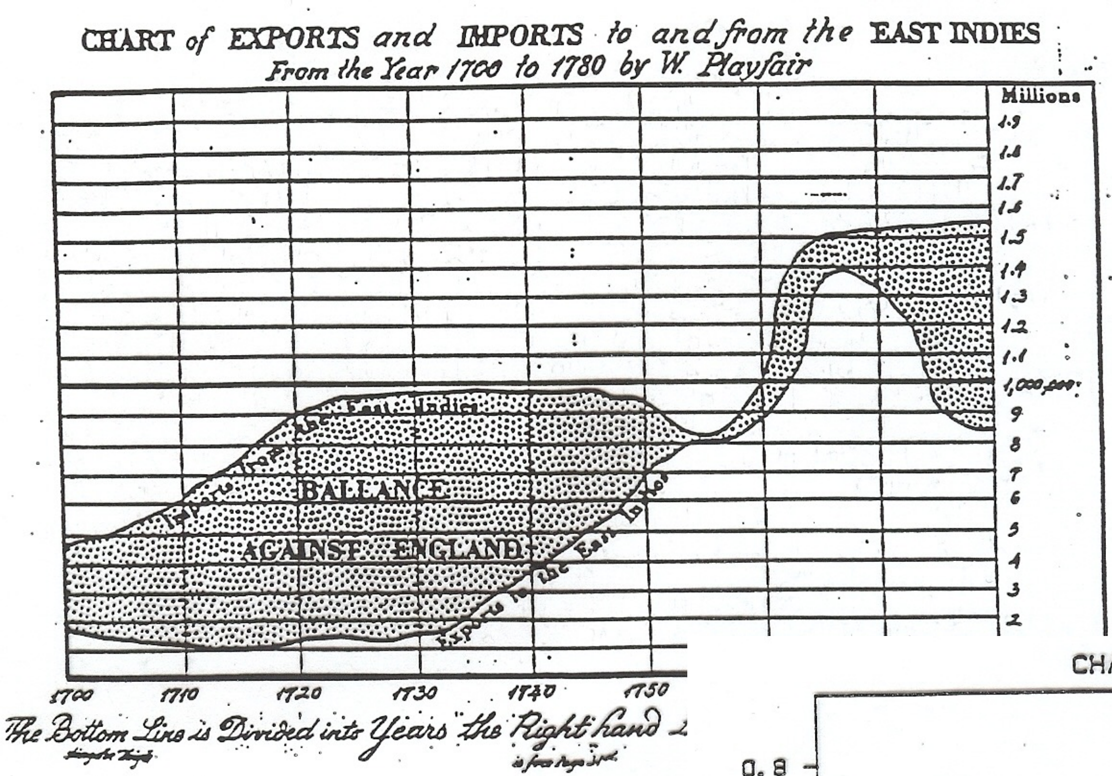


Figure 6. Curve-difference.chart alte

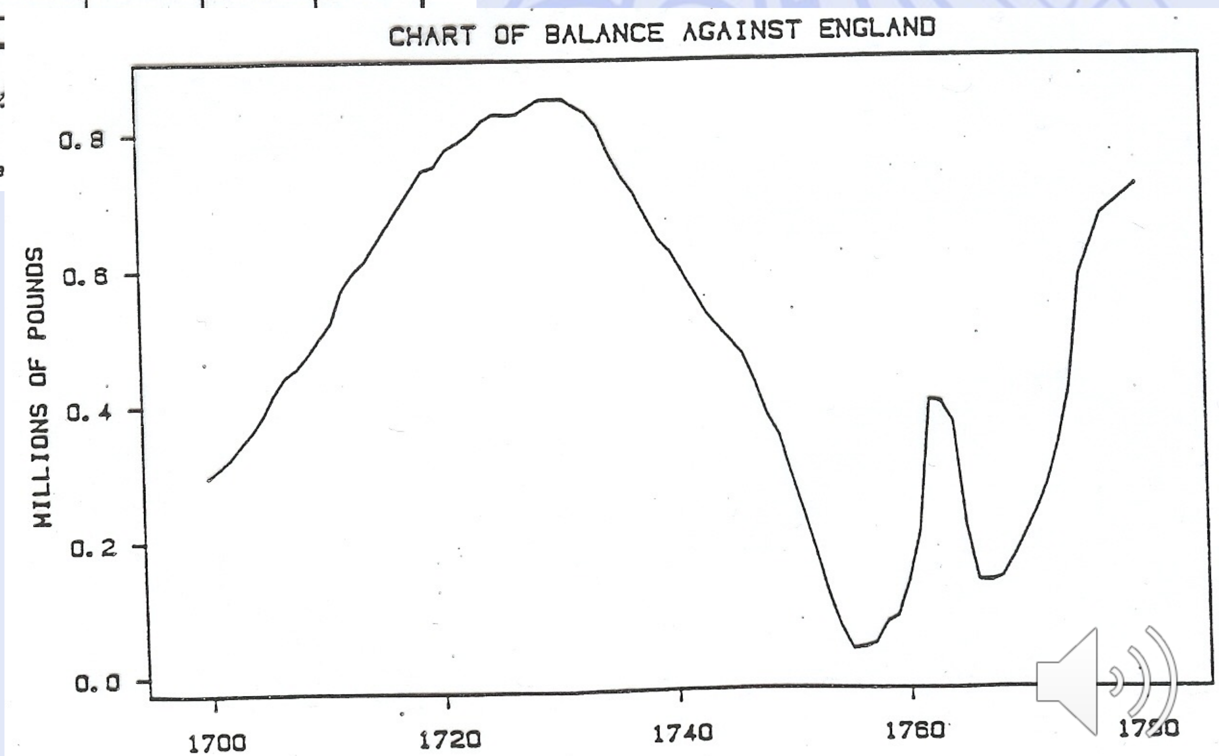


Figure 28. Playfair data.

2021/2022

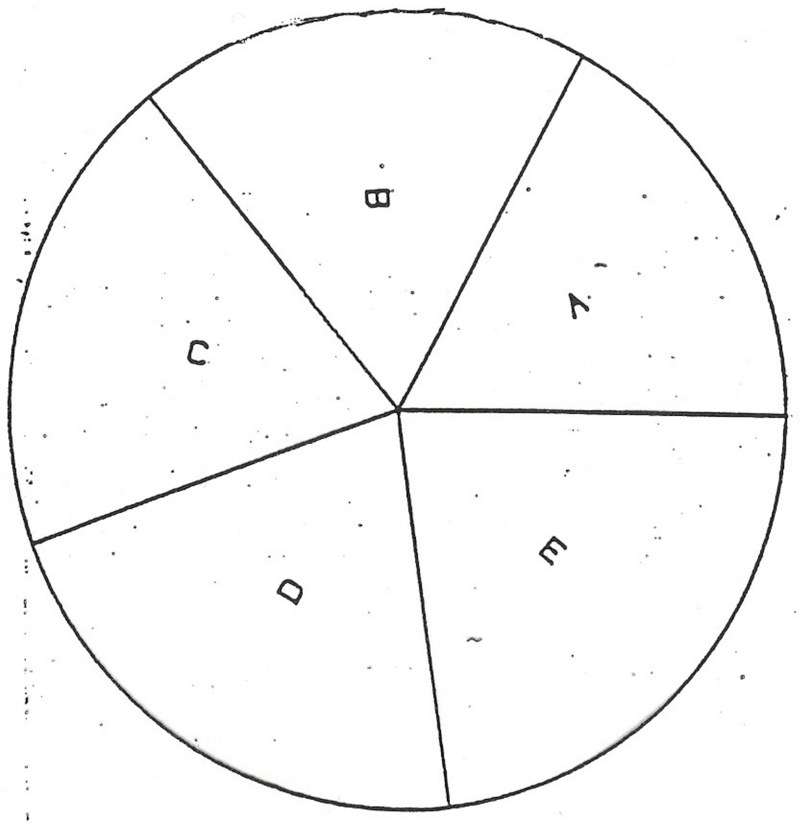


Figure 22. Pie chart.

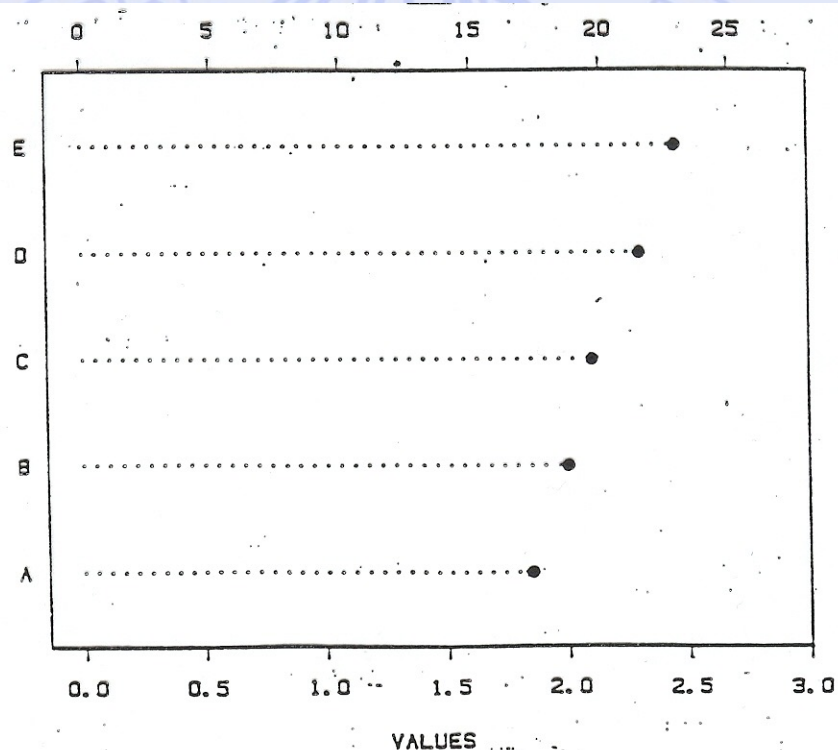


Figure 23. Dot chart.

Statistique du jour

Davis challenges Schwarzenegger to debate

Issa endorses actor-turned-politician

CNN.com, Friday, September 26, 2003 Posted: 8:14 PM EDT (0014 GMT)

LOS ANGELES, California (CNN) -- Charging that Arnold Schwarzenegger is misleading voters in the recall campaign, Democratic Gov. Gray Davis on Friday challenged the actor-turned-politician to a debate.

...

The latest poll in the recall race, by the Public Policy Institute of California, showed Bustamante with support from 28 percent of likely voters, Schwarzenegger with 26 percent and McClintock with 14 percent. With a margin of error of plus-or-minus 3 percentage points, Bustamante and Schwarzenegger were locked in a statistical tie.

...

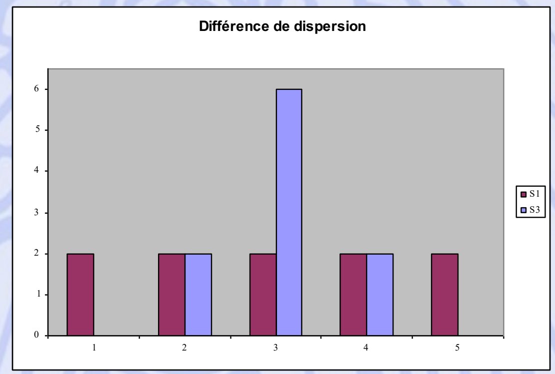
Paramètres d'une série

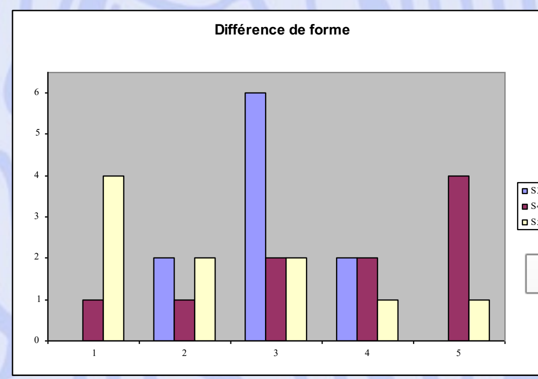
- Résumer la série en quelques grandeurs-clé !
- 3 catégories de paramètres :
  - Position
    - Valeur typique ? Valeur « centrale » ?
    - Moyenne, médiane, mode, …
  - Dispersion
    - Ecarts entre valeurs observées ?
    - Etendue, écart interquartile, variance, écart-type, …
  - Forme
    - Symétrie ? Asymétrie ?

Exemples

i	*S1*	*S2*	*S3*	*S4*	*S5*
1	1	6	2	3	3
2	2	7	3	2	4
3	2	7	3	4	2
4	1	6	2	1	5
5	3	8	3	3	3
6	4	9	3	5	1
7	5	10	4	5	1
8	3	8	3	4	2
9	4	9	3	5	1
10	5	10	4	5	1

Moyenne	3,00	8,00	3,00	3,70	2,30
Médiane	3,00	8,00	3,00	4,00	2,00
Mode	1,00	6,00	3,00	5,00	1,00
Variance	2,00	2,00	0,40	1,81	1,81
Ecart-type	1,41	1,41	0,63	1,35	1,35
Skewness	0,00	0,00	0,00	-0,80	0,80
Kurtosis	-1,33	-1,33	0,08	-0,38	-0,38





Paramètres de position

- Objectif : déterminer une valeur centrale
- ❶ Moyenne arithmétique
  - Exemple 3 :

$$\{x_i\} = \{2,0,1,1,2,3,1,1,1,0\}$$

$$\bar{x} = \frac{2+0+1+1+2+3+1+1+1+0}{10} = 1.2$$

  - Définition :

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Moyenne

- Propriétés de la moyenne arithmétique
  - Uniquement pour données numériques !
  - Fortement influencée par les valeurs extrêmes :

$$\{x_i\} = \{2,0,1,1,2,30,1,1,1,0\}$$

$$\bar{x} = \frac{2+0+1+1+2+30+1+1+1+0}{10} = 3.9$$

Moyenne

- Propriétés de la moyenne arithmétique
  - Agrégation de 2 séries :

$$\begin{bmatrix} n_1 & \bar{x}_1 \end{bmatrix} \begin{bmatrix} n_2 & \bar{x}_2 \end{bmatrix}$$

$$\Downarrow$$

$$n = n_1 + n_2 \qquad \bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n}$$

Moyenne

- Moyenne arithmétique pondérée

$$\begin{array}{cccc} x_1 & x_2 & \cdots & x_m \\ \downarrow & \downarrow & & \downarrow \\ w_1 & w_2 & \cdots & w_m \end{array}$$

$$w_k \geq 0$$

$$\sum_{k=1}^{m} w_k = 1$$

$$\bar{x}_w = \sum_{k=1}^{m} w_k x_k$$

Moyenne

- Pour une distribution observée :

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{J} n_j x_j = \sum_{j=1}^{J} f_j x_j$$

- Exemple 3 :

x_j	n_j	f_j
0	2	0.2
1	5	0.5
2	2	0.2
3	1	0.1

$$\begin{aligned}\bar{x} &= \frac{1}{10}(2\times 0 + 5\times 1 + 2\times 2 + 1\times 3) \\ &= 0,2\times 0 + 0,5\times 1 + 0,2\times 2 + 0,1\times 3 \\ &= 1,2\end{aligned}$$

Moyenne

- Pour une distribution groupée :

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{J} n_j x_{cj}$$

$(\ell_j^- ; \ell_j^+)$	x_{cj}	n_j	f_j	N_j	F_j	N_j^*	F_j^*
149.5-159.5	154.5	14	0.080	14	0.080	175	1
159.5-169.5	164.5	32	0.183	46	0.263	161	0.920
169.5-179.5	174.5	65	0.371	111	0.634	129	0.737
179.5-189.5	184.5	47	0.269	158	0.903	64	0.366
189.5-199.5	194.5	17	0.097	175	1	17	0.097
		n=175	1				

$$\bar{x} = \frac{14\times 154,5 + 32\times 164,5 + \ldots + 17\times 194,5}{175}$$

☞ Approximation !

Paramètres de position

- ❷ Médiane : $x_{1/2}$
  Valeur telle que le nombre d'observations qui la précèdent est égal au nombre d'observations qui la suivent, dans la série ordonnée.
- Exemple : $\left\{x_{(i)}\right\} = \{1,3,7,8,9,13,17\}$

$$\rightarrow x_{1/2} = 8$$

Médiane

- Pour n impair : $$x_{1/2} = x_{\left(\frac{n+1}{2}\right)}$$

- Pour n pair : (convention) $$x_{1/2} = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}$$

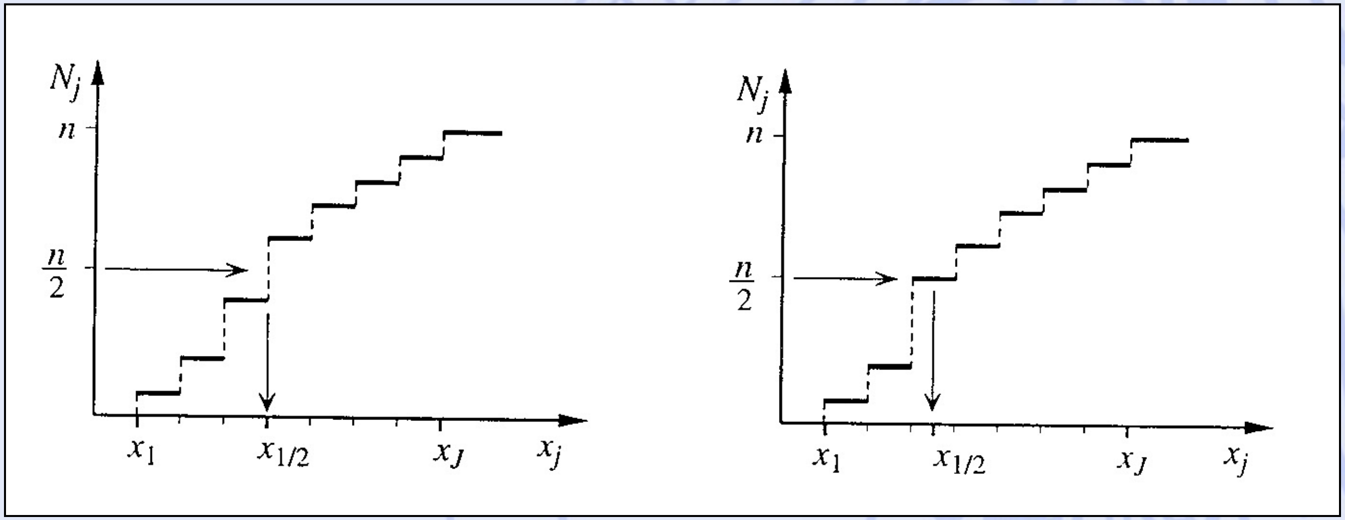
$$\left\{x_{(i)}\right\} = \{1,3,7,8,9,13\} \rightarrow x_{1/2} = \frac{7+8}{2} = 7,5$$

- Avantage de la médiane : pas influencée par les valeurs extrêmes.

Médiane

- Pour une distribution observée :

$$x_{1/2} \text{ telle que } N\left(x_{1/2}\right) = N^*\left(x_{1/2}\right)$$



$$N\left(x_{1/2}\right) = N^*\left(x_{1/2}\right) \geq \frac{n}{2}$$

Médiane

- Pour une distribution groupée :

$$x_{1/2} \text{ solution de } N\left(x_{1/2}\right) = N^{*}\left(x_{1/2}\right) = \frac{n}{2}$$

  - Classe médiane : $\left(l_m^-, l_m^+\right)$
  - Valeur approchée de la médiane :

$$x_{1/2} = l_m^- + h_m \frac{\frac{n}{2} - N_{m-1}}{n_m}$$

Médiane

- Exemple 4 :

$(\ell_j^- ; \ell_j^+)$	x_{cj}	n_j	f_j	N_j	F_j	N_j^*	F_j^*
149.5-159.5	154.5	14	0.080	14	0.080	175	1
159.5-169.5	164.5	32	0.183	46	0.263	161	0.920
169.5-179.5	174.5	65	0.371	111	0.634	129	0.737
179.5-189.5	184.5	47	0.269	158	0.903	64	0.366
189.5-199.5	194.5	17	0.097	175	1	17	0.097
		n=175	1				

$$\frac{n}{2} = \frac{175}{2} = 87,5$$

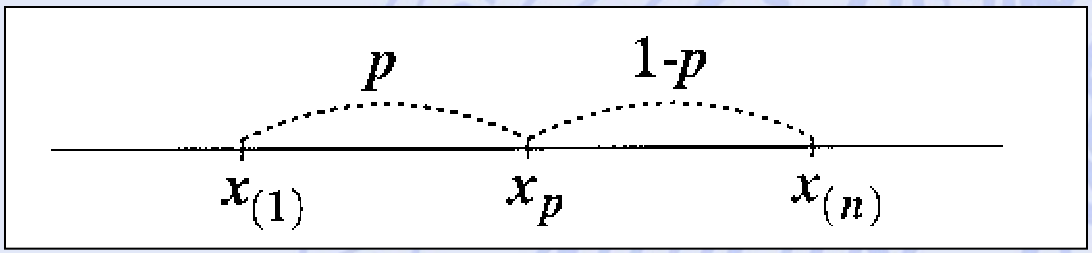
$$\left(l_m^-, l_m^+\right) = \left(169,5\ ;\ 179,5\right)$$

$$x_{1/2} = 169,5 + 10 \times \frac{87,5 - 46}{65} = 175,88$$

Paramètres de position

- ❸ Quantiles (percentiles) :

quantile d'ordre p : $x_p \quad p \in (0,1)$



$$x_p \quad \text{tel que} \quad \begin{cases} N(x_p) \geq np \\ N^*(x_p) \geq n(1-p) \end{cases}$$

Quantiles

- Quartiles :

$$p = \frac{1}{4} \qquad p = \frac{2}{4} \qquad p = \frac{3}{4}$$

$$\downarrow \qquad \downarrow \qquad \downarrow$$

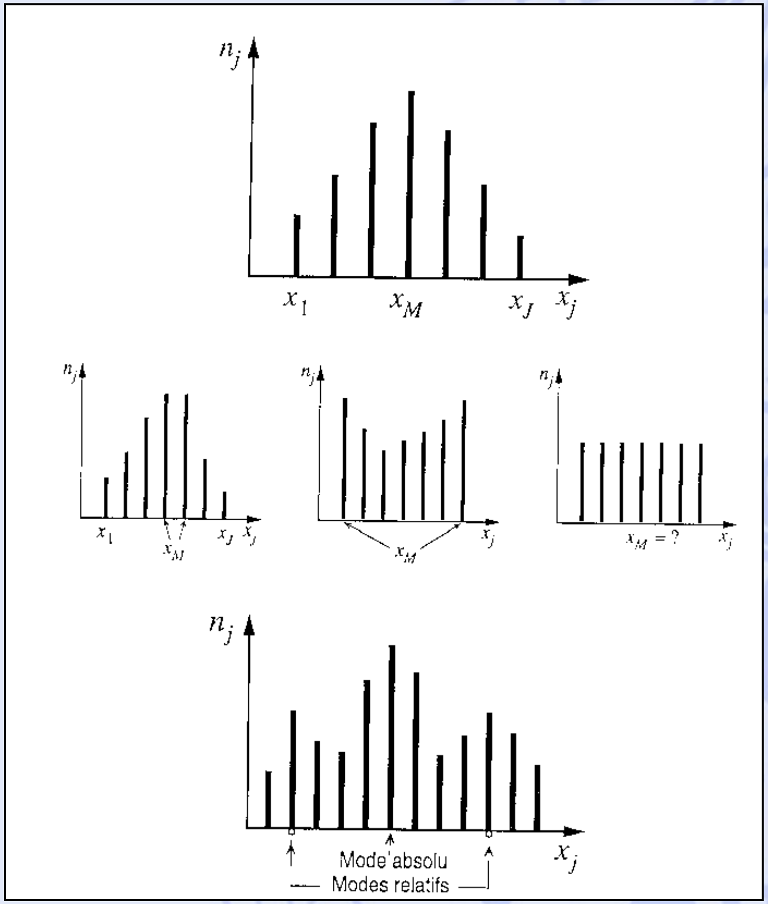
$$Q_1 = x_{1/4} \qquad Q_2 = x_{1/2} \qquad Q_3 = x_{3/4}$$

- Déciles :

$$p = \frac{1}{10} \quad p = \frac{2}{10} \quad \cdots \quad p = \frac{9}{10}$$

Paramètres de position

- ❹ Mode : Valeur la plus fréquente.



Paramètres de position

- Autres paramètres :
  - Moyenne tronquée d'ordre 1 : $\bar{x}_{t2} = \frac{1}{n-2}\sum_{i=2}^{n-1} x_{(i)}$
  - Moyenne harmonique : $H = n \Big/ \sum_{i=1}^{n} \frac{1}{x_i}$
  - Moyenne géométrique : $G = \sqrt[n]{\prod_{i=1}^{n} x_i}$

Moyenne géométrique

- pour des ratios, des taux, des pourcentages (valeurs positives uniquement).

Exemple : ventes d'une entreprise sur 5 ans :

Année	1	2	3	4	5
Ventes	100	110	132	151,8	170,02
Variation		+10%	+20%	+15%	+12%

→ Augmentation moyenne annuelle des ventes ?

$\bar{x} = 14,25\%$

$$\frac{(10+20+15+12)}{4}$$

Moyenne géométrique

$= 14,19\%$

$$\sqrt[4]{1,10 \times 1,20 \times 1,15 \times 1,12}$$

$$100 \times (1,1425)^4 = 170,38$$

$$100 \times (1,1419)^4 = 170,02$$

Moments

- Moment d'ordre r par rapport à c :

$$m_r(c) = \frac{1}{n}\sum_{i=1}^{n}(x_i - c)^r$$

- Moment par rapport à l'origine ($c = 0$) :

$$m'_r = \frac{1}{n}\sum_{i=1}^{n}x_i^r$$

- Moment centré :

$$m_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^r$$

Paramètres de dispersion

- Etendue :

$$E = x_{(n)} - x_{(1)}$$

- Ecart interquartile :

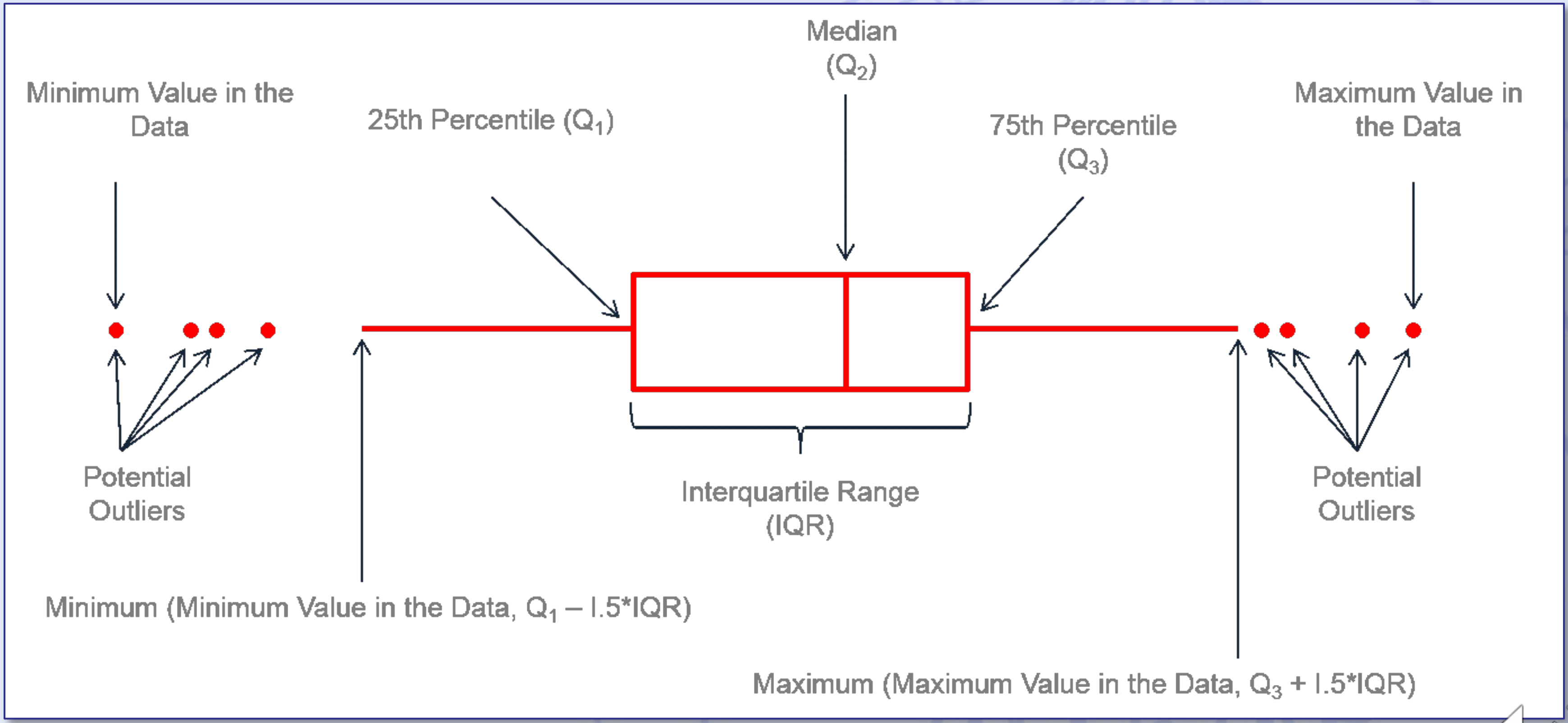
$$E_Q = x_{3/4} - x_{1/4}$$

- Ecart interdécile :

$$E_D = x_{9/10} - x_{1/10}$$

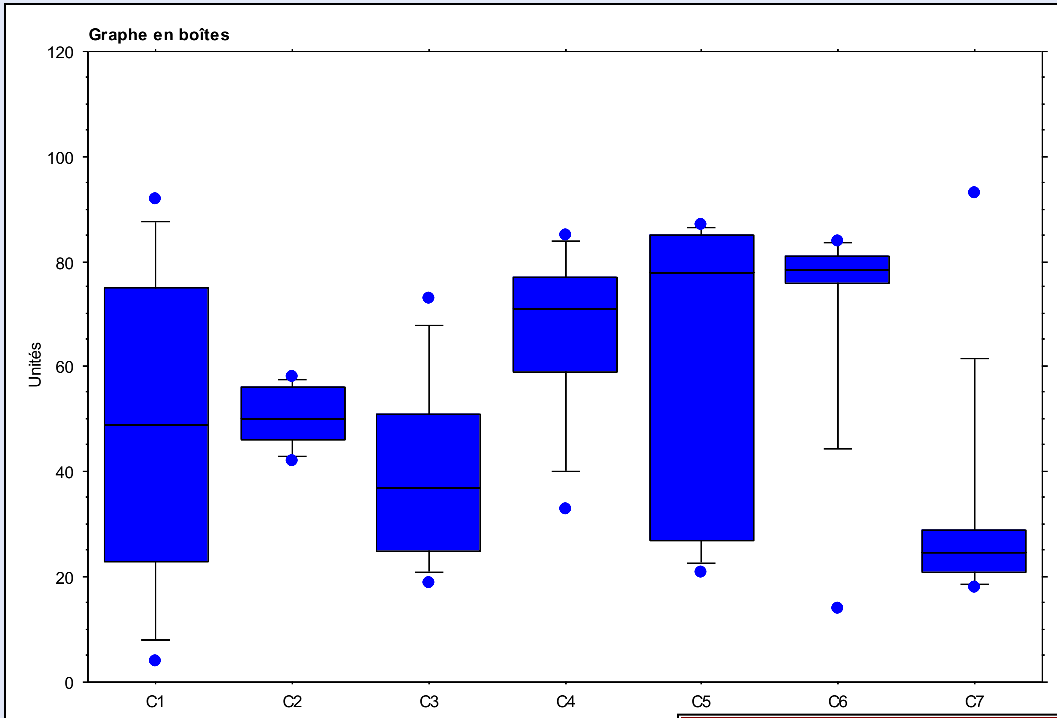
Paramètres de dispersion

- Box-Plot :



leansigmacorporation.com

Box-Plots



C1	C2	C3	C4	C5	C6	C7
4	42	19	33	21	14	18
12	44	23	47	24	75	19
23	46	25	59	27	76	21
35	47	27	67	29	77	23
46	49	31	69	77	78	24
52	51	43	73	79	79	25
67	54	48	75	83	80	27
75	56	51	77	85	81	2[illegible]
83	57	63	83	86	83	30
92	58	73	85	87	84	93

Paramètres de dispersion

- Ecart moyen absolu :

$$e_m = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - \bar{x}\right|$$

- Ecart médian absolu :

$$e_m^* = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - x_{1/2}\right|$$

Paramètres de dispersion

- Variance :

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

- Ecart-type :

$$s = \sqrt{s^2}$$

- Coefficient de variation :

$$CV = \frac{s}{\bar{x}}$$

Variance

- Agrégation de deux séries

$$\left[n_1 \quad \bar{x}_1, s_1^2\right] \quad \left[n_2 \quad \bar{x}_2, s_2^2\right]$$

$$\Downarrow$$

$$n = n_1 + n_2$$

$$s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n} + \frac{n_1\left(\bar{x}_1 - \bar{x}\right)^2 + n_2\left(\bar{x}_2 - \bar{x}\right)^2}{n}$$

Variance dans les « groupes »

Variance entre les « groupes »

Variance

- Calcul :

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2$$

- Pour une distribution observée :

$$s^2 = \frac{1}{n}\sum_{j=1}^{J} n_j (x_i - \overline{x})^2 = \frac{1}{n}\sum_{j=1}^{J} n_j x_j^2 - \overline{x}^2$$

Variance

- Exemple 3 :

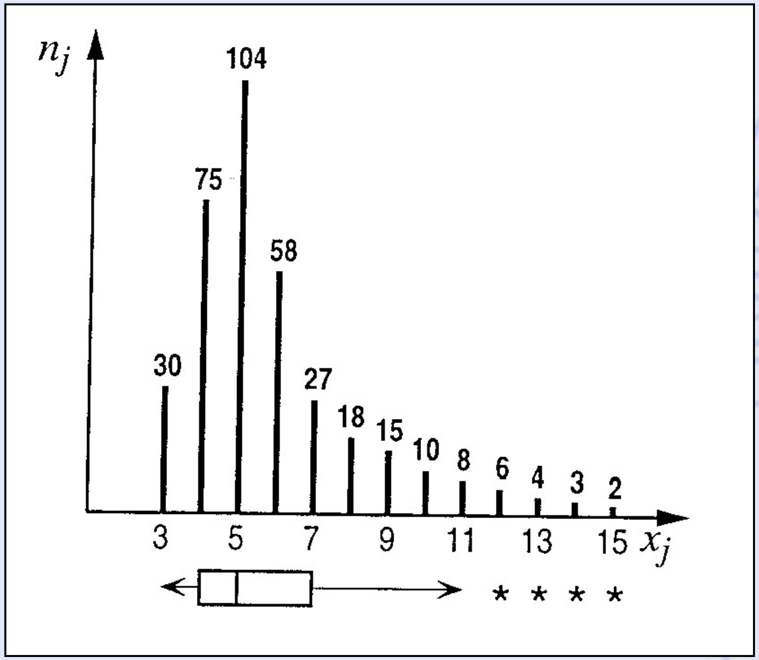
$$
\begin{aligned}
s^2 &= \frac{1}{10}\left(2\times(0-1,2)^2+5\times(1-1,2)^2+2\times(2-1,2)^2+1\times(3-1,2)^2\right)\\
&= \frac{1}{10}\left(2\times 0^2+5\times 1^2+2\times 2^2+1\times 3^2\right)-1,2^2\\
&= 0,76
\end{aligned}
$$

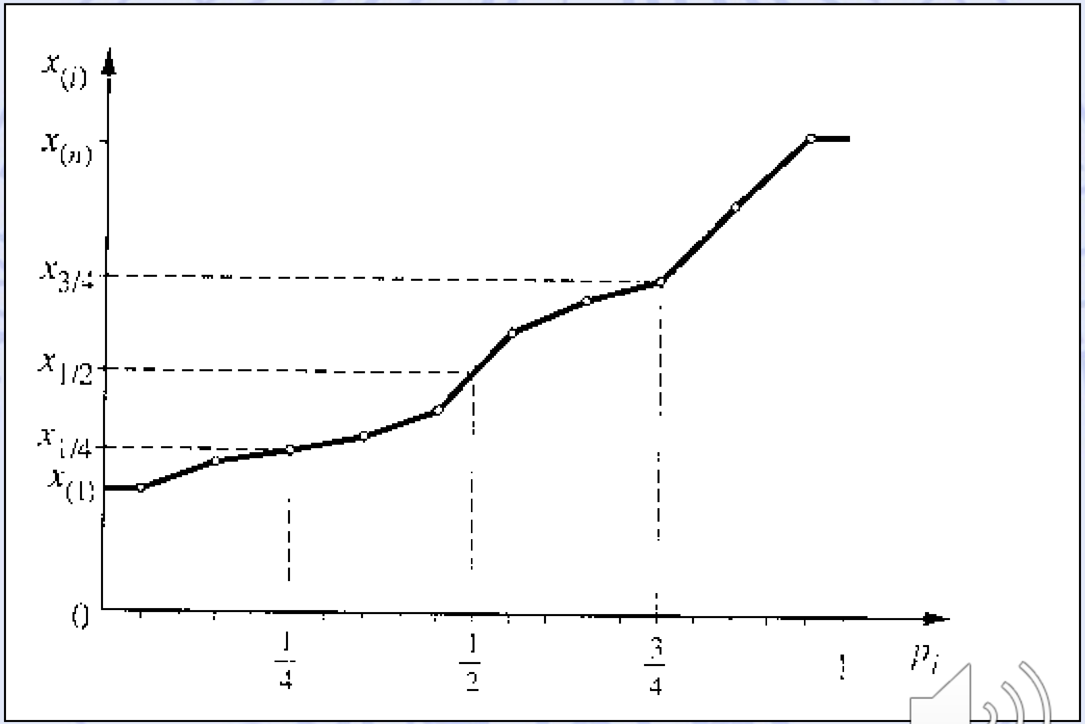
x_j	n_j	f_j
0	2	0.2
1	5	0.5
2	2	0.2
3	1	0.1

$$s=\sqrt{0,76}=0,87$$

Paramètres de forme

- Box-plot
- Graphique des quantiles

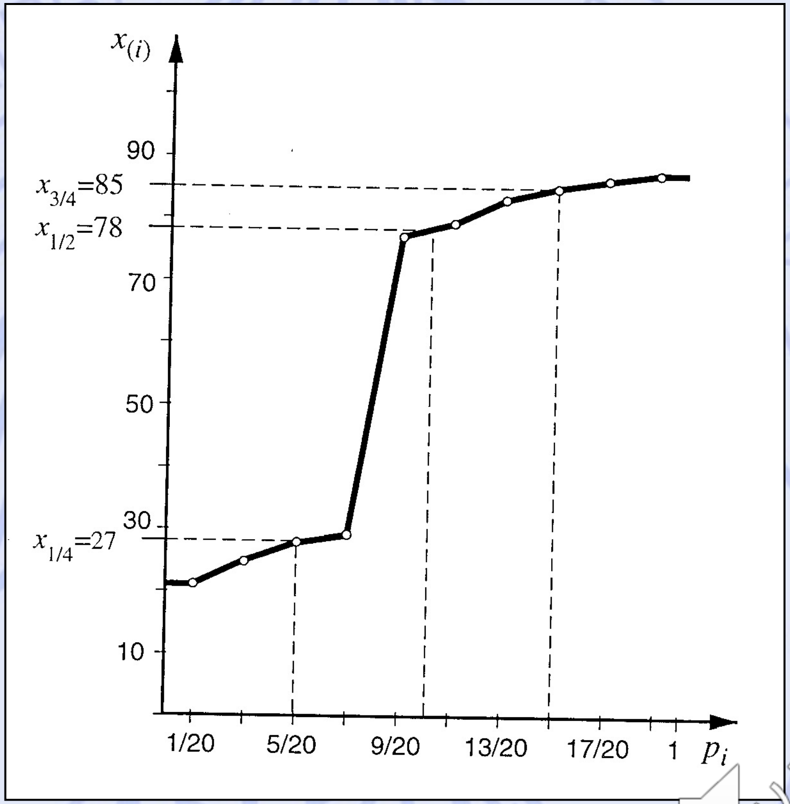




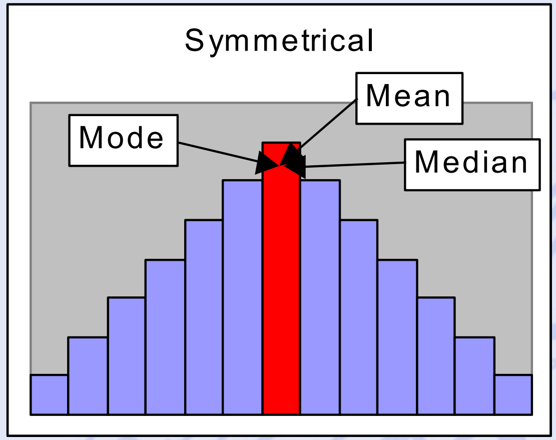
Graphique des quantiles

- Exemple C5 :

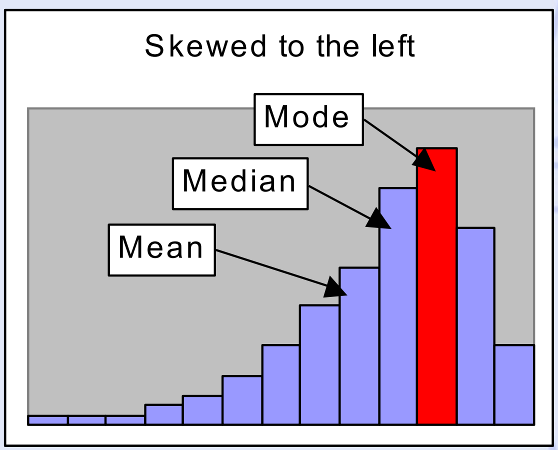
C1	C2	C3	C4	C5	C6	C7
4	42	19	33	21	14	18
12	44	23	47	24	75	19
23	46	25	59	27	76	21
35	47	27	67	29	77	23
46	49	31	69	77	78	24
52	51	43	73	79	79	25
67	54	48	75	83	80	27
75	56	51	77	85	81	29
83	57	63	83	86	83	30
92	58	73	85	87	84	93



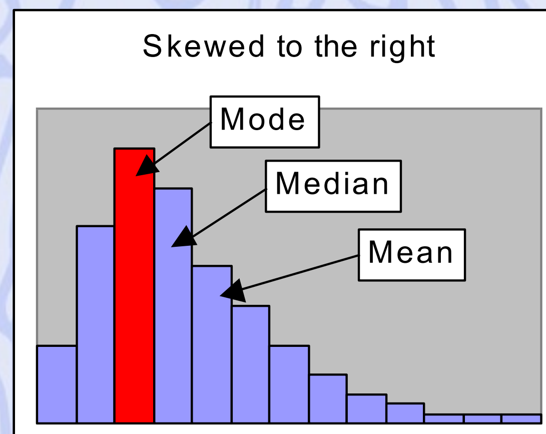
Asymétrie



Symétrie



Asymétrie à droite



Asymétrie à gauche

Coefficients d'asymétrie

- Moment centré d'ordre 3 :

$$m_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3$$

asymétrie à gauche	$m_3 > 0$
symétrie	$m_3 = 0$
asymétrie à droite	$m_3 < 0$

- Coefficient de Fisher (« skewness ») :

$$g_1 = m_3 / s^3$$

Coefficients d'asymétrie

- Coefficient empirique de Pearson :

$$S_k = \frac{\bar{x} - x_M}{s}$$

- Coefficient empirique de Yule et Kendall :

$$Y_K = \frac{x_{1/4} + x_{3/4} - 2x_{1/2}}{x_{3/4} - x_{1/4}}$$

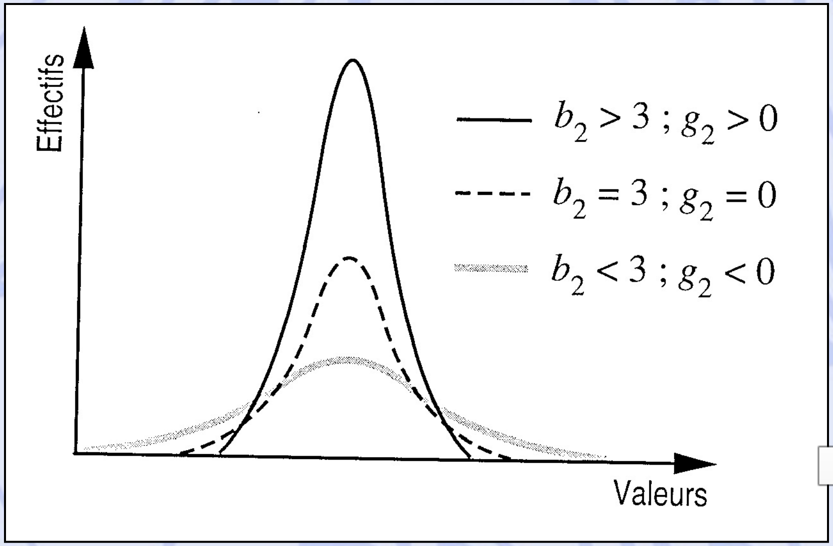
Coefficients d'aplatissement

- Coefficient de Pearson :

$$b_2 = \frac{m_4}{s^4} \quad \text{où} \quad m_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^4$$

- Coefficient de Fisher (Kurtosis) :

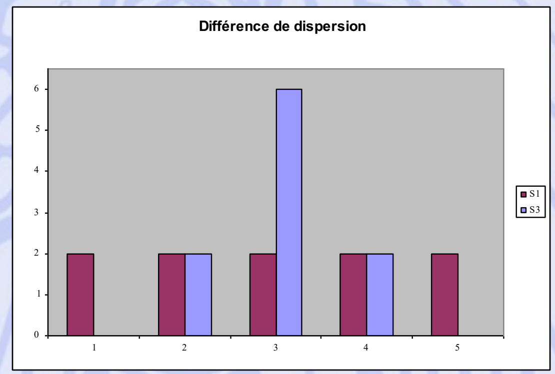
$$g_2 = b_2 - 3$$

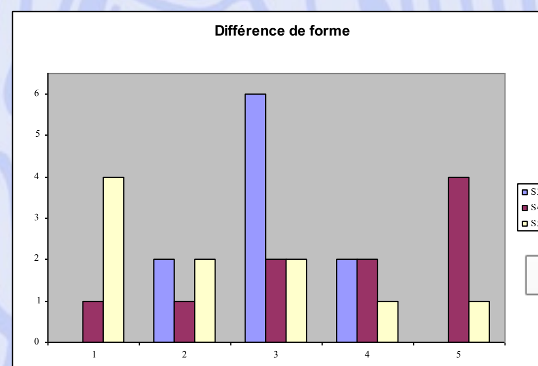


Exemples

i	*S1*	*S2*	*S3*	*S4*	*S5*
1	1	6	2	3	3
2	2	7	3	2	4
3	2	7	3	4	2
4	1	6	2	1	5
5	3	8	3	3	3
6	4	9	3	5	1
7	5	10	4	5	1
8	3	8	3	4	2
9	4	9	3	5	1
10	5	10	4	5	1

Moyenne	3,00	8,00	3,00	3,70	2,30
Médiane	3,00	8,00	3,00	4,00	2,00
Mode	~~1,00~~	~~6,00~~	3,00	5,00	1,00
Variance	2,00	2,00	0,40	1,81	1,81
Ecart-type	1,41	1,41	0,63	1,35	1,35
Skewness	0,00	0,00	0,00	-0,80	0,80
Kurtosis	-1,33	-1,33	0,08	-0,38	-0,38





Transformations de variables

- Transformation linéaire :
  - Changement d'origine et d'unité

$$x \rightarrow u = \frac{x - x_0}{d} \qquad x_0 \in R_0 \quad d \in R_0^{+}$$

  - Exemple :

$$t_{\circ C} = \frac{t_{\circ F} - 32}{1{,}8}$$

Transformations de variables

- Influence sur les paramètres de position :

$$x_i \quad \rightarrow \quad u_i = \frac{x_i - x_0}{d}$$

$$\bar{u} = \frac{\bar{x} - x_0}{d} \Leftrightarrow \bar{x} = x_0 + \bar{u}d$$

- Même transformation pour les autres paramètres de position :

$$u_{1/2} \quad u_M \quad u_p \quad \ldots$$

Transformations de variables

- Influence sur les paramètres de dispersion :
  - Changement d'origine : sur CV seulement,
  - Changement d'unité : sur tous sauf CV.

$$\left(E \quad E_Q \quad E_D \quad s^2 \quad s\right)$$

- Exemple :

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \qquad s_u^2 = \frac{1}{n}\sum_{i=1}^{n}(u_i - \overline{u})^2$$

$$\rightarrow s_u^2 = \frac{s_x^2}{d^2}$$

Transformations de variables

- Influence sur les paramètres de forme :
  - Pas d'influence sur $g_1 \quad S_k \quad Y_K \quad b_2 \quad g_2$
- Autres transformations :
  - Transformation logarithmique : $y = \log_{10} x$
  - Famille de transformations de Tukey :

$$\ldots, \frac{1}{x^2}, \frac{1}{x}, \frac{1}{\sqrt{x}}, \log x, \sqrt{x}, x, x^2, x^3, \ldots$$

  - Différence (données chronologiques) :

$$\nabla x_t = x_t - x_{t-1}$$